



TUGAS AKHIR – SS141501

***BOOSTING MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA KLASIFIKASI
DIABETES MELITUS TIPE II DI KLINIK AS-SHAFI,
SIDOARJO***

**WIDYA UMROATUN SA'DIYAH
NRP 1315 105 027**

**Dosen Pembimbing
Dr. Bambang Widjanarko Otok, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



TUGAS AKHIR – SS141501

***BOOSTING MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA KLASIFIKASI
DIABETES MELITUS TIPE II DI KLINIK AS-SHAFA,
SIDOARJO***

**WIDYA UMROATUN SA'DIYAH
NRP 1315 105 027**

**Dosen Pembimbing
Dr. Bambang Widjanarko Otok, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



FINAL PROJECT – SS141501

**BOOSTING MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) ON DIABETES
MELLITUS TYPE II CLASSIFICATIONS AT THE
AS-SHAFA CLINIC, SIDOARJO**

**WIDYA UMROATUN SA'DIYAH
NRP 1315 105 027**

**Supervisor
Dr. Bambang Widjanarko Otok, M.Si**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUTE OF TECHNOLOGY SEPULUH NOPEMBER
SURABAYA 2017**

LEMBAR PENGESAHAN

BOOSTING MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) PADA KLASIFIKASI DIABETES MELITUS TIPE II DI KLINIK AS-SHAFA, SIDOARJO

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada

Program Studi Sarjana Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

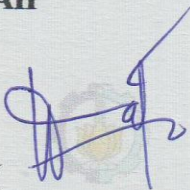
Oleh :

WIDYA UMROATUN SA'DIYAH

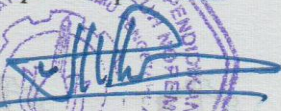
NRP. 1315 105 027

Disetujui oleh Pembimbing:

Dr. Bambang Widjanarko Otok, M.Si
NIP. 19681124 199412 1 001

()

Mengetahui,
Kepala Departemen


Dr. Suhartono
NIP. 19710929 199512 1 001


SURABAYA, JULI 2017

BOOSTING MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) PADA KLASIFIKASI DIABETES MELITUS TIPE II DI KLINIK AS-SHAFA, SIDOARJO

Nama Mahasiswa : Widya Umroatun Sa'diyah
NRP : 1315 105 027
Jurusan : Statistika FMIPA-ITS
Dosen Pembimbing : Dr. Bambang Widjanarko Otok, M.Si

Abstrak

Provinsi Jawa Timur menempati urutan ke sembilan dengan prevalensi Diabetes Melitus (DM) sebesar 6,8 dan Kota Surabaya menempati peringkat pertama prevalensi diabetes dibandingkan daerah lainnya di Jawa Timur. Tingginya angka prevalensi diabetes, menyebabkan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan membuat Program Rujuk Balik agar meningkatkan kualitas pelayanan kesehatan bagi peserta BPJS dan salah satu klinik yang mejadi rujukan peserta BPJS dan penderita DM ialah klinik As-Shafa, Sidoarjo. Oleh karena itu dilakukan analisis klasifikasi DM tipe II dengan menggunakan metode boosting MARS, dimana MARS merupakan metode yang tidak terlalu ketat dengan asumsi bentuk kurva dan boosting merupakan suatu metode yang dapat meningkatkan ketelitian dalam proses klasifikasi. Berdasarkan analisis, diketahui bahwa kombinasi model dengan $BF=10$, $MI=3$ dan $MO=10$ adalah model terbaik dan didapatkan empat variabel yang berkontribusi dalam model, yaitu variabel jenis kelamin (x_2), BMI (x_3), usia (x_1) dan aktivitas olahraga (x_5) dengan tingkat kepentingan masing-masing 100%, 60,165%, 42,682% dan 34,576%. Klasifikasi dengan metode MARS data training memiliki nilai 1-APER 0,602 dan data testing 0,529, sedangkan untuk boosting MARS data training stabil saat 250 kombinasi pengulangan dengan nilai 1-APER 0,722 dan data testing saat kombinasi pengulangan 800 dengan nilai 1-APER 0,741 yang artinya boosting MARS dapat meningkatkan ketetapan klasifikasi.

Kata Kunci : Diabetes Melitus (DM), MARS, Boosting

(Halaman ini sengaja dikosongkan)

BOOSTING MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) ON DIABETES MELLITUS TYPE II CLASSIFICATIONS AT THE AS-SHAFI CLINIC, SIDOARJO

Student Name : Widya Umroatus Sa'diyah
NRP : 1315 105 027
Department : Statistics FMIPA-ITS
Supervisor : Dr. Bambang Widjanarko Otok, M.Si

Abstract

The province of east java ranked in ninth place with 6.8 prevalence of diabetes mellitus (DM) in Indonesia and Surabaya ranked in the first rank of prevalence of diabetes in Surabaya than the other in east java. The high value of prevalence of diabetes, causing social security health agency (BPJS) make the “turning program” in order to increase the quality of health service for the BPJS member. One of the clinic that can be referred by the members of BPJS and the DM patients is As-Shafi clinic, Sidoarjo. Therefore, the classifications analysis of DM type II done by using the MARS method boosting, which MARS is a method that not too with the form and boosting is a method that might improve the carefulness in the process of classification. Based on analysis, it has seen that the combination model with BF=10, MI=3 and MO=10 is best and also obtained four variables who contributed in the model. They are type of sex (x_2), BMI (x_3), age (x_1) and sports activity (x_5) with the extent of the interest 100% , 60,165% , 42,682% and 34,576% respectively. Training data on MARS method having value 1-APER 0,602 and testing data 0,529 , while for boosting MARS training data is stable at a combination of repetition 250 with a value of 1-APER 0,722 and testing data at a combination of repetition 800 with a value of 1-APER 0,741 which means boosting mars can increase the decision of the classifications .

Keywords : Diabetes Mellitus (DM), MARS, Boosting

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul ***“Boosting Multivariate Adaptive Regression Spline (MARS) Pada Klasifikasi Diabetes Melitus Tipe II Di Klinik As-Shafa, Sidoarjo”*** ini dengan baik dan tepat waktu. terselesaikannya Tugas Akhir ini tidak terlepas dukungan dan do’a serta semangat yang diberikan dari semua pihak kepada penulis. Pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Bambang Widjanarko Otok, M.Si selaku dosen pembimbing Tugas Akhir, atas waktu yang telah diberikan dalam membimbing dan memberi masukan bagi penulis.
2. Bapak Dr. Sutikno, M.Si dan Ibu Diaz Fitra Aksioma, S.Si, M.Si selaku dosen penguji yang telah memberikan banyak saran dan masukan demi kesempurnaan Tugas Akhir ini.
3. Bapak Dr. Suhartono selaku Ketua Jurusan Statistika
4. Dr. Nurul Hidayati selaku penanggung jawab klinik As-Shafa dan Mbak Dini serta Mbak Citra selaku pegawai klinik As-Shafa, Sidoarjo
5. Seluruh Dosen, Staff, dan Karyawan Jurusan Statistika ITS yang telah membantu proses perkuliahan.
6. Orang tua penulis yang selalu memberikan motivasi dan dukungan serta tak henti-hentinya memberikan doa dalam segala kesempatan.
7. Teman-teman Lintas Jalur S-1 Statistika ITS angkatan 2015 yang telah bersedia bertukar pikiran serta diskusi.

Penulis menyadari masih banyak kekurangan dalam Laporan Tugas Akhir ini maka segala kritik dan saran demi perbaikan Laporan Tugas Akhir ini sangat dibutuhkan penulis.

Semoga laporan ini dapat bermanfaat bagi para pembaca pada umumnya dan bagi penulis pada khususnya.

Surabaya, Juli 2017

Penulis

DAFTAR ISI

| | Halaman |
|-----------------------------------------------------------|---------|
| HALAMAN JUDUL | i |
| LEMBAR PENGESAHAN | iii |
| ABSTRAK | v |
| ABSTRACT | vii |
| KATA PENGANTAR | ix |
| DAFTAR ISI | xi |
| DAFTAR GAMBAR | xiii |
| DAFTAR TABEL | xv |
| DAFTAR LAMPIRAN | xvii |
| BAB I PENDAHULUAN | |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Tujuan Penelitian | 4 |
| 1.4 Manfaat Penelitian | 4 |
| BAB II TINJAUAN PUSTAKA | |
| 2.1 Multivariate Adaptive Regression Spline (MARS) | 5 |
| 2.2 Pengujian Signifikansi Model MARS..... | 11 |
| 2.3 Probabilitas MARS Respon Biner | 11 |
| 2.4 <i>Boosting</i> | 12 |
| 2.5 Ketetapan Klasifikasi | 15 |
| 2.6 Diabetes Melitus Tipe II | 16 |
| 2.7 Faktor Risiko Terjadinya DM tipe II | 16 |
| BAB III METODOLOGI PENELITIAN | |
| 3.1 Sumber Data..... | 19 |
| 3.2 Variabel Penelitian..... | 19 |
| 3.3 Langkah analisis..... | 20 |
| BAB IV ANALISIS DAN PEMBAHASAN | |
| 4.1 Karakteristik Pasien DM tipe II di Klinik As-Shafa .. | 23 |
| 4.2 Pembentukan Model MARS | 26 |
| 4.3 Ketetapan Klasifikasi MARS | 33 |
| 4.4 Ketetapan Klasifikasi <i>Boosting</i> MARS | 34 |
| BAB V PENUTUP | |
| 5.1 Kesimpulan | 37 |
| 5.2 Saran | 38 |

| | |
|------------------------------|-----------|
| DAFTAR PUSTAKA | 39 |
| LAMPIRAN..... | 41 |
| BIODATA PENULIS | 55 |

DAFTAR GAMBAR

| | Halaman |
|----------------------------------------------------------------------------------|---------|
| Gambar 2.1 Ilustrasi <i>Boosting</i> MARS..... | 14 |
| Gambar 2.1 Kerangka Konseptual Penelitian | 17 |
| Gambar 3.1 Langkah Analisis Data | 21 |
| Gambar 4.1 Kadar Glukosa Darah (Y) Pasien DM tipe II | 23 |
| Gambar 4.2 Tekanan Darah (X_4) Pasien DM tipe II | 25 |
| Gambar 4.3 <i>Boxplot</i> variabel usia (X_2) dan BMI (X_3) | 26 |

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

| | Halaman |
|-----------------------------------------------------------------------------------------------------------|---------|
| Tabel 2.1 Tabel Klasifikasi..... | 15 |
| Tabel 2.2 Kategori Klasifikasi nilai BMI | 17 |
| Tabel 3.1 Variabel Prediktor..... | 19 |
| Tabel 3.2 Gula Darah Patokan DM | 19 |
| Tabel 4.1 Usia (X_1) Pasien DM tipe II..... | 23 |
| Tabel 4.2 Tabulasi kadar glukosa darah (Y) dan jenis kelamin (X_2) | 24 |
| Tabel 4.3 BMI (X_3) Pasien DM tipe II | 24 |
| Tabel 4.4 Tabulasi kadar glukosa darah (Y) dan senam diabetes (X_5) | 25 |
| Tabel 4.5 Hasil Pemodelan MARS | 27 |
| Tabel 4.5 Hasil Pemodelan MARS (lanjutan)..... | 28 |
| Tabel 4.6 Tingkat kepentingan variabel prediktor | 31 |
| Tabel 4.7 Hasil pengujian secara simultan..... | 32 |
| Tabel 4.8 Hasil pengujian secara parsial | 33 |
| Tabel 4.9 Klasifikasi DM tipe II data training | 33 |
| Tabel 4.10 Klasifikasi DM tipe II data testing..... | 34 |
| Tabel 4.11 Ketetapan Klasifikasi MARS | 34 |
| Tabel 4.12 Perbandingan Ketetapan Klasifikasi MARS dan <i>Boosting</i> MARS Data Training | 35 |
| Tabel 4.13 Perbandinganketetapan Klasifikasi MARS dan <i>Boosting</i> MARS Data Testing | 36 |

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

| | Halaman |
|--------------------------------------------------------------------------------------------|---------|
| Lampiran A Data Rekap Medis Pasien DM tipe II..... | 41 |
| Lampiran B <i>Output</i> Model Terbaik MARS (BF=10, MI=3 dan MO=10) | 43 |
| Lampiran C Program <i>Boosting</i> MARS | 47 |
| Lampiran D <i>Output Boosting</i> MARS (Data training dengan 600 pengulangan) | 48 |
| Lampiran E Pembuktian Pengujian Dependensi | 52 |
| Lampiran F Legalitas Data | 54 |

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hiperglikemia adalah suatu kondisi medik berupa peningkatan kadar glukosa dalam darah melebihi batas normal, yaitu kadar glukosa darah puasa tinggi jika >125 mg/dL dan merupakan salah satu tanda khas penyakit diabetes melitus (DM). Pada penelitian epidemiologi menunjukkan adanya kecenderungan peningkatan angka insidensi dan prevalensi DM tipe II di berbagai penjuru dunia. Badan Kesehatan Dunia (WHO) memprediksi adanya peningkatan jumlah penyandang DM yang menjadi salah satu ancaman kesehatan global. WHO memprediksi kenaikan jumlah penyandang DM di Indonesia dari 8,4 juta pada tahun 2000 menjadi sekitar 21,3 juta pada tahun 2030 (PERKENI, 2015). Terdapat dua kategori utama DM, yaitu DM tipe I dan DM tipe II, namun penderita DM tipe II mencapai 90% dari keseluruhan populasi penderita diabetes. DM tipe I atau *Insulin dependent* ditandai dengan kurangnya produksi insulin dan DM tipe II atau *non insulin independent* disebabkan penggunaan insulin yang kurang efektif oleh tubuh. (Depkes R.I, 2014).

Berdasarkan data hasil Riset Kesehatan Dasar (RISKEDAS) diketahui bahwa di Indonesia terdapat 10 juta orang penderita DM dan 17,9 juta orang yang beresiko menderita DM. Sementara itu, untuk Provinsi Jawa Timur menempati urutan ke sembilan dengan prevalensi 6,8 dan Kota Surabaya menempati peringkat pertama prevalensi diabetes dibandingkan daerah lainnya di Jawa Timur. Tingginya angka prevalensi diabetes di Indonesia, menyebabkan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan membuat Program Rujuk Balik, yaitu salah satu program unggulan guna meningkatkan kualitas pelayanan kesehatan bagi peserta BPJS kesehatan serta memudahkan akses pelayanan kesehatan kepada peserta penderita penyakit kronis, salah satunya ialah DM. Pasien yang merupakan peserta BPJS dan didiagnosa mengalami DM, maka akan dirujuk ke Fasilitas Kesehatan (Faskes) tingkat 1

(BPJS, 2014). Di Jawa Timur, salah satu klinik yang mejadi rujukan peserta BPJS yang juga penderita DM ialah klinik As-Shafa yang terletak di Kabupaten Sidoarjo.

Faktor risiko DM dapat dikelompokkan menjadi dua, yaitu faktor yang tidak dapat dimodifikasi dan yang dapat dimodifikasi. Faktor risiko yang yang tidak dapat dimodifikasi adalah ras dan etnik, umur jenis kelamin, riwayat keluarga diabetes melitus, riwayat melahirkan bayi dengan berat badan lebih dari 4000 gram, dan riwayat lahir dengan berat badan lahir rendah (kurang dari 2500 gram). Sedangkan faktor risiko yang dapat dimodifikasi erat kaitannya dengan perilaku hidup yang kurang sehat, yaitu berat badan lebih yang menyebabkan obesitas abdominal/sentral, kurangnya aktivitas fisik, hipertensi, diet tidak sehat (Depkes R.I, 2014).

Beberapa kajian mengenai DM telah dilakukan, diantaranya Ningtyas (2013) yang menganalisis kualitas hidup pasien diabetes melitus tipe II yang hasilnya variabel lama menderita diabetes, komplikasi diabetes dan tingkat pendidikan berpengaruh signifikan, sedangkan Fernanda (2012) menganalisis tentang *Boosting neural network* dan *Boosting CART* pada klasifikasi diabetes melitus tipe II yang hasilnya metode *Boosting CART* menghasilkan akurasi yang lebih tinggi dibandingkan *Boosting neural network* dengan variabel yang digunakan ialah riwayat keturunan, umur, jenis kelamin, obesitas, pola makan dan aktivitas olahraga. Penelitian mengenai diabetes mellitus tipe II tersebut mengindikasikan bahwa banyak sekali faktor yang mempengaruhi Diabetes Mellitus tipe II. Oleh karena itu, untuk menjelaskan pola hubungan antar variabel respon dengan variabel prediktor dapat menggunakan kurva regresi. Pendekatan kurva regresi yang sering digunakan adalah pendekatan regresi parametrik, dimana diasumsikan bentuk kurva linier. Namun, tidak semua pola hubungan antar variabel dapat didekati dengan pendekatan parametrik, karena tidak adanya suatu informasi mengenai bentuk hubungan antara variabel respon dan variabel prediktor. Jika asumsi model parametrik tidak terpenuhi, maka kurva regresi dapat diduga dengan

pendekatan model regresi nonparametrik. Beberapa model regresi Nonparametrik yang banyak digunakan, antara lain adalah *kernel*, *k-nearest neighbor* (Hardle, 1990) dan *multivariate regression spline* (MARS).

MARS merupakan pendekatan untuk regresi nonparametrik yang dikembangkan oleh Friedman (1990) dengan menggunakan fungsi *spline* untuk menduga model. Pertimbangan penggunaan metode MARS untuk pemodelan ini, didasarkan pada ketidakjelasan pola hubungan antara variabel respon variabel prediktor. MARS merupakan metode yang tidak terlalu ketat dengan asumsi bentuk kurva, sehingga memberikan fleksibilitas yang tinggi, bahkan pada yang melibatkan banyak interaksi. Beberapa penelitian dengan metode MARS telah banyak dilakukan, antara lain Hidayanti (2014) tentang *Boosting MARS binary response* untuk klasifikasi kemiskinan yang hasilnya tingkat akurasi klasifikasi dengan model MARS lebih tinggi daripada *Boosting MARS* pada sampel kecil, sehingga saran yang diberikan adalah penambahan jumlah sampel, serta Adityaningrum (2016) tentang *propensity score matching* menggunakan pendekatan MARS pada kasus DM yang hasilnya variabel yang berpengaruh adalah jenis kelamin, dislipidemia (gula darah) dan obesitas.

Data yang digunakan dalam penelitian ini adalah data rekap medik pasien DM di klinik As-Shafa, Sidoarjo yang terdiagnosis memiliki penyakit DM tipe II dan merupakan pengguna layanan BPJS dengan variabel yang digunakan ialah usia, jenis kelamin, *body mass index* (BMI), tekanan darah, dan aktivitas olahraga (senam diabetes), sedangkan variabel responnya adalah kadar glukosa darah puasa pasien. Semua variabel tersebut secara bersama-sama dapat dibentuk menjadi model MARS, sehingga dapat dilakukan klasifikasi untuk mengetahui ketetapan klasifikasinya. Dalam implementasi untuk klasifikasi, terkadang menghasilkan ketetapan klasifikasi yang kurang memuaskan. Hal ini dapat dilihat berdasarkan tingkat akurasi. Oleh karena itu, dapat digunakan metode resampling dalam penyusunan modelnya untuk menurunkan tingkat kesalahan klasifikasi. Salah satu metode resam-

pling yang dapat digunakan adalah *boosting*. *Boosting* merupakan suatu metode yang dapat meningkatkan ketelitian dalam proses klasifikasi dengan cara membangkitkan kombinasi dari suatu model, karena data yang mengalami kesalahan klasifikasi akan mengalami penambahan bobot sehingga diharapkan metode *Boosting* MARS ini memberikan akurasi klasifikasi yang lebih baik daripada metode MARS.

1.2 Rumusan Masalah

Banyaknya faktor yang mempengaruhi risiko DM tipe II, maka dilakukan analisis klasifikasi untuk mengetahui ketetapan klasifikasi dari faktor-faktor risiko DM. Salah satu metode klasifikasi ialah metode MARS, namun hasil dari klasifikasi MARS terkadang kurang memuaskan, sehingga dilakukan resampling dengan menggunakan metode *Boosting* MARS.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui variabel yang berpengaruh signifikan pada klasifikasi dengan metode MARS terhadap data penderita DM tipe II serta membandingkan hasil tingkat ketetapan klasifikasi metode MARS dengan *Boosting* MARS.

1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat, antara lain :

1. Bagi bidang pendidikan, penelitian ini diharapkan dapat memberikan wawasan keilmuan dalam penerapan metode MARS dan *Boosting* MARS.
2. Bagi bidang kesehatan, penelitian ini dapat memberikan dapat informasi mengenai faktor-faktor risiko DM tipe II.

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka dalam penelitian ini merupakan penjelasan dari metode yang digunakan antara lain Multivariate Adaptive Regression Spline (MARS), *Boosting* MARS, ketetapan klasifikasi dalam pemilihan model terbaik, pengertian diabetes melitus serta faktor-faktor terjadinya DM.

2.1 Multivariate Adaptive Regression Spline (MARS)

Model MARS digunakan untuk mengatasi permasalahan data berdimensi tinggi (Hestie, Tibshirani, & Friedman, 2008). Selain itu, model MARS juga menghasilkan klasifikasi variabel respon yang akurat, serta menghasilkan model yang kontinu dalam knot berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil (Friedman, 1990). GCV merupakan salah satu metode untuk mendapatkan knot optimum. MARS dikembangkan oleh Friedman (1990) untuk pendekatan model regresi nonparametrik antara variabel respon beberapa variabel prediktor pada *piecewise* regresi. *Piecewise* regresi merupakan regresi yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih daripada polinomial biasa. Apabila suatu garis regresi tidak bisa menjelaskan keseluruhan data, maka beberapa garis regresi digunakan untuk menjelaskan seluruh data yang ada dari variabel yang independen. Tempat perubahan pola atau garis regresi itulah yang dinamakan knot. Knot ini merupakan akhir dari sebuah garis regresi (region) dan awal dari sebuah garis regresi yang lain. Pada setiap knot diharapkan adanya kontinuitas dari basis fungsi antar satu region dengan region yang lain. Basis fungsi merupakan suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan prediktor. Friedman (1990) menyarankan jumlah *basis maximum* (BF) fungsi adalah dua sampai empat kali jumlah variabel prediktor, sedangkan jumlah *maximum interaction* (MI) adalah satu, dua, atau tiga dengan pertimbangan jika lebih dari tiga akan menghasilkan model yang sangat kompleks

serta jarak minimum antar knot atau *minimum observation* (MO) sebanyak nol, satu, dua, tiga, lima dan sepuluh.

MARS merupakan pengembangan dari pendekatan *recursive partitioning regression* (RPR). Akan tetapi, pendekatan *recursive partitioning regression* (RPR) masih memiliki kelemahan dimana model yang dihasilkan tidak kontinu pada knot, sehingga model MARS digunakan untuk mengatasi kelemahan *Recursive Partitioning Regression* (RPR), yaitu menghasilkan model yang kontinu pada knot dan dapat mengidentifikasi adanya fungsi linier dan aditif. Perbaikan yang dilakukan untuk mengatasi keterbatasan RPR, antara lain menghasilkan basis fungsi pada persamaan (2.1).

$$B_m^q(\mathbf{x}) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+^q \quad (2.1)$$

K_m adalah banyaknya interaksi pada fungsi basis m , t_{km} merupakan knot dari variabel prediktor $x_{v(k,m)}$ dan s_{km} nilainya +1 jika knotnya terletak di kanan dan -1 jika knotnya terletak di kiri subregion. v merupakan banyaknya variabel prediktor dan m merupakan banyaknya basis fungsi serta k merupakan banyaknya interaksi. Modifikasi Friedman dalam mengatasi kelemahan RPR menghasilkan persamaan (2.2)

$$\hat{f}(\mathbf{x}) = \gamma_0 + \sum_{m=1}^M \gamma_m \prod_{k=1}^{K_m} [S_{K_m} \cdot (x_{v(k,m)} - t_{K_m})]_+ \quad (2.2)$$

dengan fungsi,

$$(x_{v(k,m)} - t_{km})_+ = \begin{cases} (x_{v(k,m)} - t_{km}), & x_{v(k,m)} - t_{km} > 0 \\ 0, & x_{v(k,m)} - t_{km} \leq 0 \end{cases}$$

Dimana a_0 adalah koefisien konstanta dari basis fungsi B_0 . Estimasi untuk $\{\gamma_m\}_{m=0}^M$ ditentukan menggunakan metode kuadrat terkecil, sehingga dalam bentuk matriks jadi persamaan (2.3).

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (2.3)$$

Dimana :

$$\mathbf{y} = (y_1, \dots, y_n)^T ; \quad \boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_M)^T ; \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T \text{ dan}$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} (s_{k1} (x_{1v(k,1)} - t_{k1}))_+ & \cdots & \prod_{k=1}^{K_m} (s_{km} (x_{1v(k,M)} - t_{kM}))_+ \\ 1 & \prod_{k=1}^{K_1} (s_{k1} (x_{2v(k,1)} - t_{k1}))_+ & \cdots & \prod_{k=1}^{K_m} (s_{km} (x_{2v(k,M)} - t_{kM}))_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} (s_{k1} (x_{nv(k,1)} - t_{k1}))_+ & \cdots & \prod_{k=1}^{K_m} (s_{km} (x_{nv(k,M)} - t_{kM}))_+ \end{bmatrix}$$

Pada persamaan dalam bentuk matriks di atas, \mathbf{y} merupakan vektor variabel respon yang berukuran $(n \times 1)$ dan \mathbf{B} merupakan matriks basis fungsi yang berukuran $(n \times (M+1))$, sedangkan $\boldsymbol{\gamma}$ adalah vektor koefisien regresi yang berukuran $((M+1) \times 1)$ dan $\boldsymbol{\varepsilon}$ adalah vektor *error* yang berukuran $(n \times 1)$.

Sum Square Error (SSE) diperoleh dengan cara mengkuadratkan persamaan berikut.

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{B}\boldsymbol{\gamma}$$

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}) \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\gamma}^T \mathbf{B}^T \mathbf{y} + \boldsymbol{\gamma}^T \mathbf{B}^T \mathbf{B} \boldsymbol{\gamma} \end{aligned}$$

Perhitungan nilai \mathbf{B} dengan melakukan diferensial SSE

$$\frac{\partial \text{SSE}}{\partial \boldsymbol{\gamma}} = -2\mathbf{B}^T \mathbf{y} + 2\mathbf{B}^T \mathbf{B} \boldsymbol{\gamma} = 0$$

$$\mathbf{B}^T \mathbf{y} = \mathbf{B}^T \mathbf{B} \boldsymbol{\gamma}$$

$$\hat{\boldsymbol{\gamma}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{y}$$

Penjabaran dari persamaan (2.2) dapat dilihat pada persamaan (2.4).

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \gamma_0 + \sum_{m=1}^M \gamma_m (s_{1m} (x_{v(1,m)} - t_{(1,m)}))_+ \\ &\quad + \sum_{m=1}^M \gamma_m (s_{1m} (x_{v(1,m)} - t_{(1,m)}))_+ (s_{2m} (x_{v(2,m)} - t_{(2,m)}))_+ \\ &\quad + \sum_{m=1}^M \gamma_m (s_{1m} (x_{v(1,m)} - t_{(1,m)}))_+ (s_{2m} (x_{v(2,m)} - t_{(2,m)}))_+ (s_{3m} (x_{v(3,m)} - t_{(3,m)}))_+ + \dots \end{aligned} \quad (2.4)$$

Misal diambil $BF=3$ dan $MI=2$, maka persamaan (2.4) dapat ditulis sebagai berikut.

$$\begin{aligned}\hat{f}(\mathbf{x}) &= \gamma_0 + \sum_{m=1}^6 \gamma_m \prod_{k=1}^2 [S_{Km} \cdot (x_{v(k,m)} - t_{Km})]_+ \\ \hat{f}(\mathbf{x}) &= \gamma_0 + \gamma_1 (S_{11} \cdot (x_{v(1,1)} - t_{(1,1)})) \\ &\quad + \gamma_2 (S_{12} \cdot (x_{v(1,2)} - t_{(1,2)})) + \gamma_3 (S_{13} \cdot (x_{v(1,3)} - t_{(1,3)})) \\ &\quad + \gamma_1 [(S_{11} \cdot (x_{v(1,1)} - t_{(1,1)})) (S_{21} \cdot (x_{v(2,1)} - t_{(2,1)}))] \\ &\quad + \gamma_2 [(S_{12} \cdot (x_{v(1,2)} - t_{(1,2)})) (S_{22} \cdot (x_{v(2,2)} - t_{(2,2)}))] \\ &\quad + \gamma_3 [(S_{13} \cdot (x_{v(1,3)} - t_{(1,3)})) (S_{23} \cdot (x_{v(2,3)} - t_{(2,3)}))]\end{aligned}$$

Sehingga persamaan (2.4) dapat ditulis menjadi persamaan (2.5).

$$\hat{f}(\mathbf{x}) = \gamma_0 + f_i(x_i) + f_{ij}(x_i, x_j) + f_{ijk}(x_i, x_j, x_k) + \dots + f_{ijk\dots}(x_i, x_j, x_k, \dots) \quad (2.5)$$

Diberikan $V(m) = \{v(k, m)\}_1^{K_m}$ sehingga masing-masing fungsi pada persamaan (2.5) dapat dijabarkan sebagai berikut :

$$\begin{aligned}f_i(x_i) &= \sum_{m=1}^M \gamma_m B_m(x_i) \\ f_{ij}(x_i, x_j) &= \sum_{m=1}^M \gamma_m B_m(x_i, x_j) \\ f_{ijk}(x_i, x_j, x_k) &= \sum_{m=1}^M \gamma_m B_m(x_i, x_j, x_k)\end{aligned}$$

Persamaan (2.5) menunjukkan suku kedua meliputi penjumlahan semua basis fungsi untuk satu variabel, suku ketiga meliputi penjumlahan semua basis fungsi untuk interaksi antar dua variabel, suku keempat meliputi penjumlahan semua basis fungsi untuk interaksi antara tiga variabel dan seterusnya. Interpretasi model MARS melalui persamaan tersebut adalah mempresentasikan variabel yang masuk dalam model, baik untuk satu variabel maupun interaksi antar variabel.

Untuk mempermudah interpretasi model MARS, maka persamaan model MARS pada (2.3) dapat disederhanakan pada persamaan (2.7)

$$\hat{f}(\mathbf{x}) = \gamma_0 + \gamma_1 BF_1 + \gamma_2 BF_2 + \dots + \gamma_M BF_M \quad (2.6)$$

Dengan $\hat{f}(\mathbf{x})$ merupakan variabel respon, a_0 adalah konstanta, γ_M adalah koefisien untuk basis fungsi ke- M , dan BF_M adalah basis fungsi ke- M .

Dalam pembentukan model MARS, langkah awal adalah menentukan titik-titik perubahan pola perilaku data atau yang disebut dengan titik knot. Pemilihan knot pada MARS menggunakan algoritma *forward* dan *backward*. Algoritma MARS, khususnya tahap *forward* digunakan untuk mendapatkan subregion-subregion agar dapat menentukan basis fungsi. Penentuan titik knot dan koefisien sangat penting agar mendapatkan model terbaik. Untuk memenuhi konsep parsimoni (model sederhana) dilakukan tahap *backward*, yaitu mengeluarkan basis fungsi yang berkontribusi kecil terhadap nilai dugaan respon.

Ukuran kontribusi yang digunakan pada tahap ini menggunakan kriteria GCV. GCV digunakan karena memiliki sifat optimal asimtotik (Wahba, 1990). Metode GCV secara umum seperti persamaan (2.8)

$$GCV(M) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2 \left/ \left[1 - \frac{C(M)}{n} \right]^2 \right. \quad (2.7)$$

Keterangan:

M : jumlah fungsi basis yang ditentukan

$C(M)$: jumlah parameter dalam model = $\text{Trace} \left(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \right) + 1$

\mathbf{B} : matriks basis fungsi

n : jumlah data

y_i : nilai variabel respon pada pengamatan ke- i

$\hat{f}(x_i)$: nilai taksiran variabel respon pada pengamatan ke- i

Kriteria lain yang digunakan ialah nilai MSE (Eubank, 1999) dan Koefisien Determinasi (R^2) (Draper, 1992) yang dihitung menggunakan persamaan (2.8) dan (2.9)

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{f}(x_i))^2}{n - M - 1} \quad (2.8)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{f}(x_i) - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.9)$$

Algoritma pembentukan model MARS terdiri dari 2 tahapan (Permatasari, 2013).

- 1) Tahap *forward* dilakukan untuk mendapatkan fungsi dengan jumlah basis maksimum, yaitu diawali dengan menginisiasi, kemudian dilanjutkan pada tahap menduga koefisien konstanta dari basis fungsi B_0 . Misalkan terdapat M basis fungsi $B_0, B_1(x), \dots, B_M(x)$ maka basis fungsi yang digunakan :

$$B_1(x) = \prod_{k=1}^{K_1} (s_{k1} (x_{v(k,1)} - t_{k1}))_+$$

$$B_2(x) = \prod_{k=1}^{K_2} (s_{k2} (x_{v(k,2)} - t_{k2}))_+$$

Dengan $B_m(x)$ adalah basis fungsi $x_{v(k,m)}$ adalah peubah yang terdapat dalam basis fungsi $B_m(x)$ dan t_{km} adalah posisi titik knot. Dimana $t_{km} \in \{x_{v(k,m)}\}$. Penambahan basis fungsi dilanjutkan hingga M basis fungsi maksimum dan penentuannya menggunakan kriteria GCV.

- 2) Tahap *backward* akan dipilih satu fungsi basis (kecuali B_0) dan mengeluarkan basis tersebut jika kontribusinya kecil, kemudian proses *backward* akan berlanjut hingga tidak ada fungsi basis yang dapat dikeluarkan. Misalkan fungsi basis $B_m(x)$ dengan $m=0,1,2,\dots,M$, untuk menduga koefisien regresi dari γ_m dengan menggunakan metode kuadrat terkecil, $\gamma = (\gamma_1, \dots, \gamma_M)^T$ dengan $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ serta \mathbf{B} merupakan matriks komponen $B_m(x)$.

2.2 Pengujian Signifikansi model MARS

Apabila telah ditemukan model MARS terbaik, maka dilakukan pengujian untuk mengecek signifikansi parameter untuk mengevaluasi kecocokan model. Pengujian dilakukan dengan menguji koefisien regresi secara simultan maupun secara parsial.

1. Pengujian regresi simultan

a. Perumusan Hipotesis

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_M = 0 \text{ (model tidak signifikan)}$$

$$H_1 : \text{minimal terdapat satu } \gamma_m \neq 0 ; m = 1, 2, \dots, M \\ \text{(model signifikan)}$$

b. Taraf signifikan :

c. Statistik uji (α) : 5%

$$F_{hitung} = \frac{\sum_{i=1}^n \frac{(\hat{y}_i - \bar{y})^2}{M}}{\sum_{i=1}^n \frac{(\hat{y}_i - \bar{y})^2}{n - M - 1}} \quad (2.10)$$

d. Daerah kritis : Tolak H_0 jika nilai $F_{hitung} > F_{\alpha(M; n-M-1)}$

2. Pengujian regresi parsial

a. Perumusan Hipotesis

$$H_0 : \gamma_m = 0 \text{ (koefisien } \gamma_m \text{ tidak berpengaruh terhadap model)}$$

$$H_1 : \gamma_m \neq 0 \text{ untuk setiap } m, \text{ dimana } m = 1, 2, \dots, M \\ \text{(koefisien } \gamma_m \text{ berpengaruh terhadap model)}$$

b. Taraf signifikan (α) : 5%

c. Statistik uji :

$$t_{hitung} = \frac{\hat{\gamma}_m}{Se(\hat{\gamma}_m)} \quad (2.11)$$

dengan

$$Se(\hat{\gamma}_m) = \sqrt{Var(\hat{\gamma}_m)}$$

d. Daerah kritis : Tolak H_0 jika nilai $|t_{hitung}| > t_{(\alpha/2, n-M)}$

2.3 Probabilitas MARS Respon Biner

Pada model MARS, masalah klasifikasi didasarkan pada pendekatan analisis regresi logistik. Jika variabel respon terdiri dari dua nilai, maka dikatakan sebagai regresi respon biner,

sehingga dapat digunakan model probabilitas pada persamaan (2.12) (Otok, 2010).

$$P(Y = 1 | X = x) = \hat{\pi}(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}}$$

$$P(Y = 0 | X = x) = (1 - \hat{\pi}(x)) = \frac{1}{1 + e^{\hat{f}(x)}} \quad (2.12)$$

Variabel respon Y merupakan variabel respon biner (0 dan 1) dengan p adalah banyaknya variabel prediktor $x(x_1, x_2, \dots, x_p)$.

2.4 Boosting

Boosting adalah suatu metode yang dapat meningkatkan ketelitian dalam proses klasifikasi dengan cara membangkitkan kombinasi dari suatu model, t etapi hasil klasifikasi yang dipilih adalah model yang memiliki *misclassification* terkecil. Konsep metode *boosting* adalah memperbaiki ketetapan klasifikasi dari suatu pemodelan yang sering disebut dengan *weak learner*. Secara umum *boosting* terjadi dalam iterasi. Pada setiap iterasi dilakukan perubahan nilai bobot. Data-data yang mengalami kesalahan klasifikasi akan mengalami penambahan bobot. Oleh karena itu, *weak learner* pada iterasi selanjutnya akan lebih terfokus pada data-data yang mengalami kesalahan klasifikasi oleh *weak learner* sebelumnya (Hestie, Tibshirani, & Friedman, 2008). Algoritma yang digunakan sebagai berikut (Wu & Kumar, 2008).

Input

1. Dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$
2. Algoritma *weak learning* L
3. Jumlah replikasi R

Proses

1. Inisiasi bobot sampel data $D_1(i) = 1/n$
2. Untuk $r = 1, \dots, R$
 - a. Lakukan prediksi h_r untuk dataset D yang telah diboboti D_r :
 $h_r = L(D, D_r)$

- b. Hitung bobot D_r dari data yang salah diklasifikasikan ε_r
- c. Hitung bobot baru $\omega_r = 1/2 \ln\left(\frac{1-\varepsilon_r}{\varepsilon_r}\right)$
- d. Perbarui bobot sampel data

$$D_{r+1} = \begin{cases} \frac{D_r(i)}{Z_r} \times \exp(-\omega_r) & , h_r(x_i) = y_i \\ \frac{D_r(i)}{Z_r} \times \exp(\omega_r) & , h_r(x_i) \neq y_i \end{cases}$$

Dimana Z_r merupakan faktor normalisasi agar $\sum_{i=1}^m D_{r+1}(i) = 1$

3. Selesai

Output : Prediksi klasifikasi akhir $H(x) = \text{sign}\left(\sum_{r=1}^R \omega_r h_r(x)\right)$

Algoritma diatas digunakan untuk metode *boosting*, namun untuk metode *boosting* MARS digunakan algoritma sebagai berikut.

Input

1. Dataset $D = \{(BF_{.1}, y_1), (BF_{.2}, y_2), \dots, (BF_{.n}, y_n)\}$
2. Algoritma *weak learning* L
3. Jumlah replikasi R

Proses

1. Inisiasi bobot sampel data $D_1(i) = 1/n$
2. Untuk $r = 1, \dots, R$
 - a. Lakukan prediksi h_r untuk dataset D yang telah diboboti D_r :
 $h_r = L(D, D_r)$
 - b. Hitung bobot D_r dari data yang salah diklasifikasikan ε_r
 - c. Hitung bobot baru $\omega_r = 1/2 \ln\left(\frac{1-\varepsilon_r}{\varepsilon_r}\right)$
 - d. Perbarui bobot sampel data

$$D_{r+1} = \begin{cases} \frac{D_r(i)}{Z_r} \times \exp(-\omega_r) & , h_r(x_i) = y_i \\ \frac{D_r(i)}{Z_r} \times \exp(\omega_r) & , h_r(x_i) \neq y_i \end{cases}$$

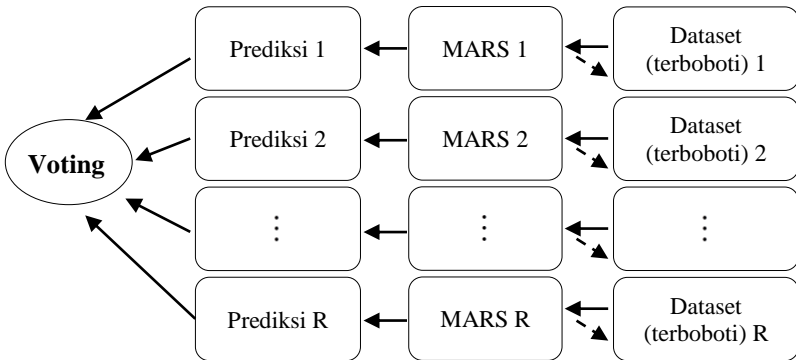
Dimana Z_r merupakan faktor normalisasi agar $\sum_{i=1}^m D_{r+1}(i) = 1$

4. Selesai

Output : Prediksi klasifikasi akhir $H(x) = \text{sign}\left(\sum_{r=1}^R \omega_r h_r(x)\right)$

Keterangan :

- Setiap training data memiliki bobot distribusi awal $D_1(i)$
- m adalah jumlah seluruh data
- ω_r selalu bernilai positif, yaitu merupakan tingkat *confidence* dari prediksi
- D_r bernilai real
- i adalah id data
- Weak learner* merupakan suatu fungsi yang mengklasifikasikan data yang telah diboboti.
- Sign* ialah tanda dari penjumlahan data, dimana bernilai 1 jika hasil $H(x)$ positif dan bernilai -1 jika hasil $H(x)$ negatif



Gambar 2.1 Ilustrasi *Boosting MARS*

Banyaknya kombinasi pengulangan pada *boosting* adalah 50,100,250 dan 500 kombinasi, namun pada penelitian lain dilakukan penambahan jumlah kombinasi hingga data stabil. Data stabil jika nilai ketetapan klasifikasi data sama untuk setiap kombinasi (Permatasari, 2013).

2.5 Ketetapan Klasifikasi

Perhitungan klasifikasi dilakukan untuk melihat seberapa besar pengelompokan sekumpulan data untuk digolongkan dengan tepat pada kelompoknya. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan klasifikasi yang kecil. Salah satu ukuran evaluasi akurasi klasifikasi ialah dengan menghitung nilai 1-APER (*apparent error rate*). APER merupakan sejumlah observasi dataset yang salah diklasifikasikan oleh sebuah fungsi klasifikasi (Johnson & Winchern, 1992).

Tabel 2.1 Tabel Klasifikasi

| Kelas prediksi | Kelas Sebenarnya | |
|----------------|------------------------|------------------------|
| | $y_1=-1$ | $y_2=1$ |
| $h_1=-1$ | TN (True Negative) | FN (False Negative) |
| $h_2=1$ | FP (False Positive) | TP (True Positive) |

dimana :

TN = Jumlah h_1 dari klasifikasi yang tepat diklasifikasikan ke y_1

FP = Jumlah h_2 dari klasifikasi yang salah diklasifikasikan ke y_1

FN = Jumlah h_1 dari klasifikasi yang salah diklasifikasikan ke y_2

TP = Jumlah h_2 dari klasifikasi yang tepat diklasifikasikan ke y_2

Maka perhitungan akurasi dapat menggunakan persamaan (2.13)

$$1 - APER = \frac{(TP + TN)}{(TN + TP + FN + FP)} \quad (2.13)$$

Adapun menurut Agresti (2007), pada tabel klasifikasi perlu dipertimbangkan juga nilai *sensitivity* dan *specificity*. *Sensitivity* menggambarkan akurasi pada sampel kelas positif, sedangkan *specificity* menggambarkan akurasi pada sampel kelas negatif. Ukuran akurasi klasifikasi yang juga mempertimbangkan *sensitivity* dan *specificity* adalah G-means.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (2.14)$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (2.15)$$

$$G - \text{means} = \sqrt{\text{sensitivity} + \text{specificity}} \quad (2.16)$$

G-means dapat menggambarkan bagaimana sebuah metode klasifikasi mampu mengukur *sensitivity* dan *specificity*. Semakin besar nilai G-means menunjukkan metode klasifikasi tersebut mampu memprediksi data di tiap kelas dengan baik.

2.6 Diabetes Melitus Tipe II

Penderita DM tipe II mencapai 90% dari keseluruhan populasi penderita diabetes faktor genetik serta lingkungan memberikan pengaruh yang cukup besar dalam menyebabkan terjadinya DM tipe II, selain itu juga terjadinya obesitas, diet tinggi lemak dan rendah serat, serta kurang gerak badan (Depkes RI, 2005). Diabetes melitus (DM) biasanya ditandai dengan tingginya kadar gula darah, oleh karena itu diperlukan pemeriksaan kadar glukosa darah secara berkala untuk *skrining* dan diagnosis DM.

Glukosa merupakan karbohidrat terpenting yang kebanyakan diserap ke dalam aliran darah sebagai glukosa. Glukosa adalah bahan bakar utama dalam jaringan tubuh serta berfungsi untuk menghasilkan energi. Kadar glukosa darah sangat erat kaitannya dengan penyakit DM. Peningkatan kadar glukosa darah puasa ≥ 125 yang disertai penurunan berat badan, *polyuria* (sering buang air kecil), *polydipsia* (sering haus), *polifagia* (banyak makan/mudah lapar) sudah cukup untuk menegakkan diagnosis DM (Amir, Wungouw, & Pangemanan, 2015). Selain itu sering pula muncul keluhan penglihatan kabur, koordinasi gerak tubuh terganggu, kesemutan pada tangan atau kaki, timbul gatal-gatal yang seringkali sangat mengganggu (*pruritus*).

2.7 Faktor Risiko Terjadinya DM tipe II

Berdasarkan penelitian Handayani (2003) membuktikan bahwa usia ≥ 45 tahun memiliki pengaruh terhadap kejadian DM tipe II. Orang yang berusia ≥ 45 tahun beresiko terkena DM tipe II sebesar 7,5 kali dibandingkan dengan mereka yang berusia < 45

tahun. Faktor yang disebabkan perilaku erat hubungannya dengan obesitas dan salah satu cara untuk mengetahui kenormalan berat badan ialah dengan menghitung *body mass index* (BMI).

$$\text{BMI} = \frac{\text{berat badan (kg)}}{\text{tinggi badan}^2 (\text{m})} \quad (2.17)$$

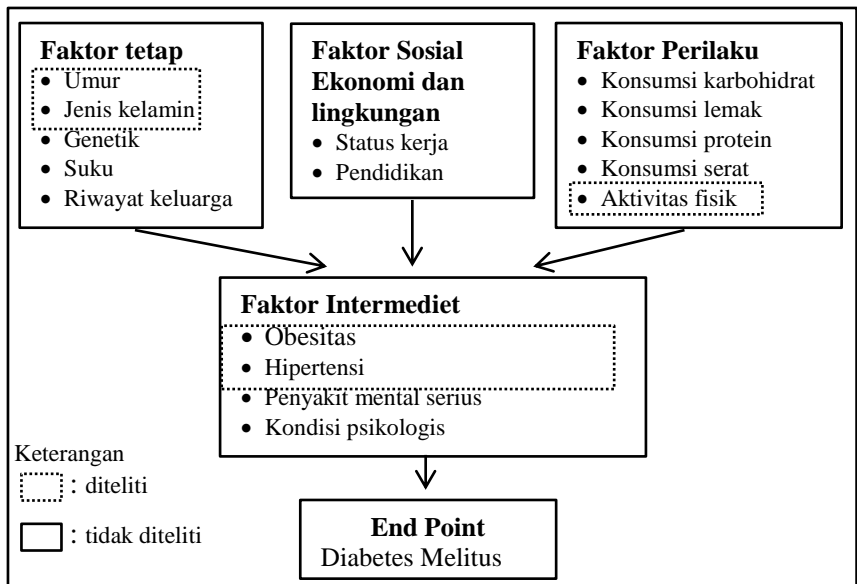
dimana kategori pengklasifikasiannya ialah sebagai berikut.

Tabel 2.2 Kategori Klasifikasi nilai BMI

| Kategori | Nilai BMI | |
|-----------|-----------|-----------|
| | Laki-laki | Perempuan |
| Kurus | <17 | <18 |
| Normal | 17-23 | 18 -25 |
| Kegemukan | 23-27 | 25-27 |
| Obesitas | >27 | >27 |

Sumber : Departemen Kesehatan RI

dan kerangka konseptual yang digunakan dalam penelitian ini didasarkan pada penelitian Garnita (2012) sebagai berikut.



Gambar 2.2 Kerangka Konseptual Penelitian

(Halaman Ini Sengaja Dikosongkan)

BAB III METODOLOGI PENELITIAN

Pada bab ini dibahas mengenai metodologi penelitian yang meliputi sumber data, variabel penelitian, langkah analisis dan diagram alir penelitian.

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa rekam medik pasien diabetes melitus tipe II di klinik As-Shafa, Sidoarjo pada bulan Maret 2017 dengan jumlah pasien sebanyak 344 pasien.

3.2 Variabel Penelitian

Variabel yang digunakan dalam analisis merujuk pada penelitian sebelumnya yang dilakukan oleh Fernanda (2012) dan Erniati (2013) pada Tabel 3.1

Tabel 3.1 Variabel Prediktor

| Simbol Variabel | Kategori | Skala data |
|------------------------------------------------|-----------------------------------|------------|
| X ₁ Usia | - | Rasio |
| X ₂ Jenis Kelamin | 1= Laki-laki 0= Perempuan | Nominal |
| X ₃ <i>Body mass index</i> (BMI) | - | Rasio |
| X ₄ <i>Tekanan darah</i> (Sistolik) | - | Rasio |
| X ₅ Aktivitas Olahraga | 1= Aktif 0= Tidak aktif/Kurang | Nominal |

dengan variabel respon (y) dalam penelitian ini adalah kadar glukosa darah penderita DM tipe II, jika 110-125 mg/dL dikategorikan kadar glukosa darah normal yang dikoding 0 dan jika ≥ 126 mg/dL dikategorikan kadar glukosa darah tinggi yang dikoding 1. Penentuan tingkat kadar glukosa darah tersebut menurut WHO dapat dilihat pada Tabel 3.2

Tabel 3.2 Gula Darah Patokan Diagnosis DM

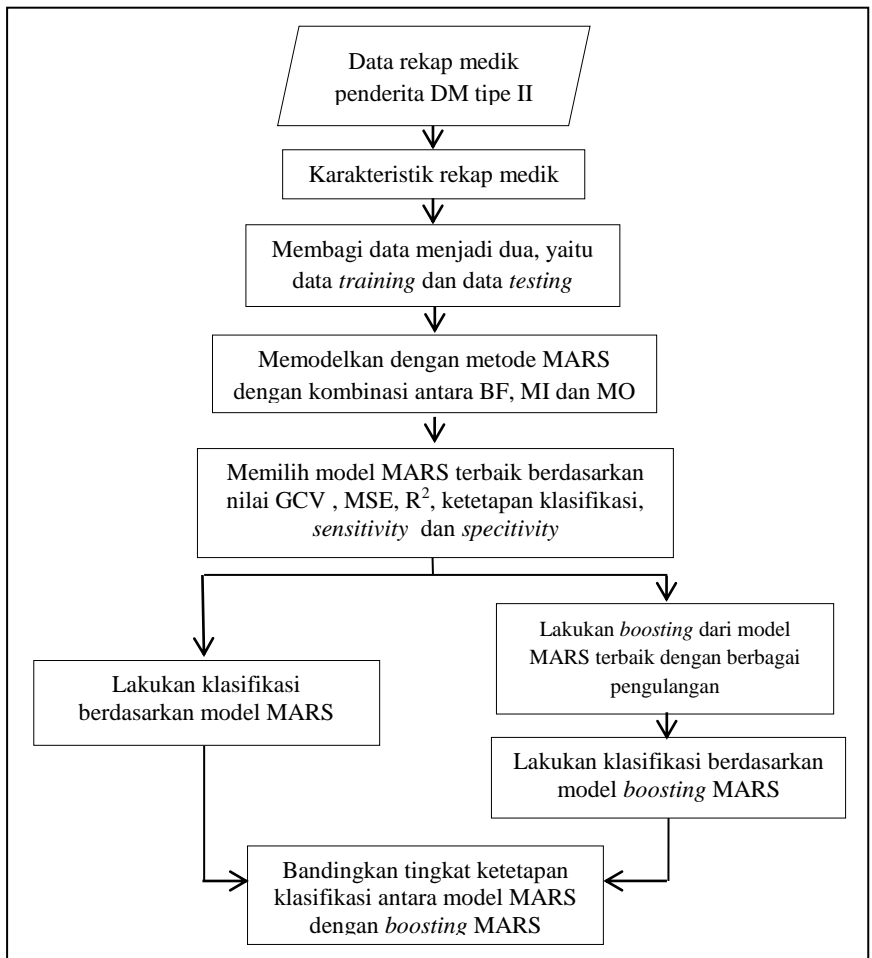
| Kadar Glukosa Darah | Bukan DM | Belum pasti DM (Normal) | DM (Tinggi) |
|----------------------------------|----------|-------------------------|-------------|
| Kadar Glukos darah puasa (mg/dL) | < 90 | 100 - 125 | ≥ 126 |

3.3 Langkah Analisis

Sebelum dilakukan analisis dengan menggunakan klasifikasi MARS, maka dilakukan analisis deskriptif untuk mengetahui karakteristik dari data pasien yang menderita diabetes Melitus tipe II. Selanjutnya dilakukan analisis klasifikasi sebagai berikut :

1. Membagi data menjadi dua, yaitu data *training* sebanyak 75% atau 259 data dan *testing* 25% atau sebanyak 85 data
2. Menyusun pemodelan untuk data pasien diabetes tipe II di klinik As-Shafa, Sidoarjo berdasarkan data training dengan menggunakan metode MARS
 - a. Memodelkan model MARS dengan langkah melakukan kombinasi antara :
 1. Maksimum basis fungsi (BF), yaitu 12,18, 24
 2. Jumlah maksimum interaksi (MI), yaitu 1, 2, 3
 3. Jarak minimum antar knot / minimum observasi (MO), yaitu 0, 1, 2, 3, 5 dan 10
 - b. Mendapatkan model MARS terbaik berdasarkan nilai GCV dan MSE terkecil serta nilai R^2 dan akurasi (1-APER) terbesar.
3. Mengklasifikasikan data training dan testing berdasarkan model MARS terbaik yang diperoleh pada langkah 2 serta menghitung ketetapan klasifikasinya.
4. Berdasarkan hasil model terbaik pada langkah (2), variabel respon yang harus digunakan dalam *boosting* MARS harus dirubah berdasarkan kategorinya, yaitu kadar glukosa darah normal kadar dan glukosa darah tinggi, sedangkan variabel prediktor yang digunakan ialah perhitungan basis fungsi (BF). Selanjutnya dilakukan analisis sebagai berikut.
 - a. Melakukan *boosting* dari model MARS dengan 50,100, 250, 500, 600, 700, 800, 900 dan 1000 pengulangan.
 - b. Melakukan klasifikasi dari data aktual dengan data prediksi masing-masing kombinasi pengulangan hasil *boosting* MARS pada langkah 4a untuk mengetahui ketetapan klasifikasi dari hasil *boosting* MARS.
5. Bandingkan tingkat ketetapan klasifikasi antara model MARS dengan *boosting* MARS

Tahapan analisis data selengkapnya disajikan pada Gambar 3.1.



Gambar 3.1 Langkah Analisis Data

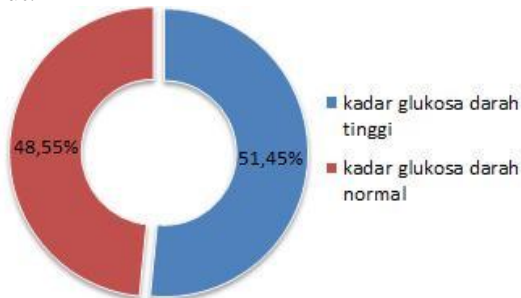
(Halaman Ini Sengaja Dikosongkan)

BAB IV ANALISIS DAN PEMBAHASAN

Bab ini menjelaskan hasil analisis dan pembahasan tentang diabetes melitus tipe II di Klinik As-Shafa, Sidoarjo. Analisis yang dilakukan meliputi karakteristik data, pemilihan model terbaik menggunakan MARS, *Boosting* MARS serta klasifikasi dan ketetapan klasifikasinya.

4.1 Karakteristik Pasien DM tipe II di Klinik As-Shafa

Untuk mengetahui karakteristik data, maka dilakukan analisis deskriptif untuk mengetahui gambaran umum dari data yang digunakan dalam suatu penelitian. Dalam penelitian ini terdapat lima variabel prediktor dan satu variabel respon dengan analisis sebagai berikut.



Gambar 4.1 Kadar Glukosa Darah (Y) Pasien DM tipe II

Gambar 4.1 menunjukkan kadar glukosa darah Pasien DM tipe II di klinik As-Shafa. Pengkategorian kadar glukosa darah tinggi jika 110-126 mg/dL dan kadar glukosa normal jika tekanan berada 110-125 mg/dL. Pada Gambar 4.1 terlihat bahwa sebesar 51,45% atau 177 pasien yang memiliki kadar glukosa darah tinggi dan 48,55% lainnya atau sebanyak 167 pasien memiliki kadar glukosa darah normal.

Tabel 4.1 Usia (X_1) Pasien DM tipe II

| Variabel | Rata-rata | Varsians | Minimum | Maksimum |
|----------|-----------|----------|----------|-----------|
| Usia | 63 tahun | 86 | 29 tahun | 112 tahun |

Selanjutnya ialah karakteristik untuk variabel usia pasien (X_1) DM tipe II yang dapat dilihat pada Tabel 4.1 menunjukkan rata-rata usia pasien DM tipe II di klinik As-Shafa ialah 63 tahun, dimana di usia tersebut tubuh sudah lemah dalam memproses makanan sesuai dengan keperluannya dan usia minimum pasien adalah 29 tahun serta maksimum usia pasien 112 tahun, sehingga nilai varians sangat tinggi, yaitu 86.

Tabel 4.2 Tabulasi kadar glukosa darah (Y) dan jenis kelamin (X_2)

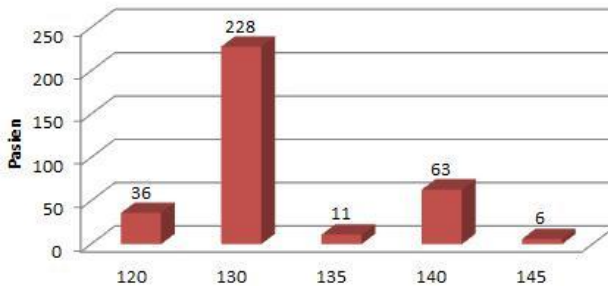
| Pasien DM Tipe II | Perempuan | Laki-laki |
|----------------------------|-----------|-----------|
| Kadar Glukosa darah normal | 91 | 76 |
| Kadar Glukosa darah tinggi | 111 | 66 |
| Total | 202 | 142 |

Tabel 4.2 menunjukkan tabulasi antara kadar glukosa darah (Y) dan jenis kelamin (X_2) diketahui bahwa dari 202 pasien perempuan, 91 pasien memiliki kadar glukosa darah normal dan 111 lainnya memiliki kadar glukosa darah tinggi, sedangkan untuk jenis kelamin laki-laki, diketahui bahwa dari 142 pasien, 76 memiliki kadar glukosa darah normal dan 66 lainnya memiliki kadar glukosa darah tinggi.

Tabel 4.3 BMI (X_3) Pasien DM tipe II

| Jenis Kelamin | Rata-rata | Varians | Maksimum | Minimum |
|---------------|-----------|---------|----------|---------|
| Laki-laki | 24,38 | 19,63 | 45,92 | 15,98 |
| Perempuan | 24,48 | 26,39 | 45,92 | 15,98 |

Faktor terjadinya DM yang disebabkan oleh perilaku erat hubungannya dengan terjadinya obesitas dan salah satu cara untuk mengetahui kenormalan berat badan, yaitu dengan menghitung *body mass index* (BMI) yang dapat dilihat pada Tabel 4.3. Rata-rata BMI dari pasien DM untuk laki-laki dan perempuan hampir sama yaitu 24, namun untuk laki-laki BMI tersebut dapat dikatakan mengalami kegemukan (obesitas) berdasarkan Tabel 2.2, sedangkan untuk perempuan masih dikatakan normal. Untuk nilai maksimum dan minimum BMI sama untuk laki-laki dan perempuan, namun nilai BMI untuk perempuan memiliki varians yang tinggi, yaitu 26,39



Gambar 4.2 Tekanan Darah (X_4) Pasien DM tipe II

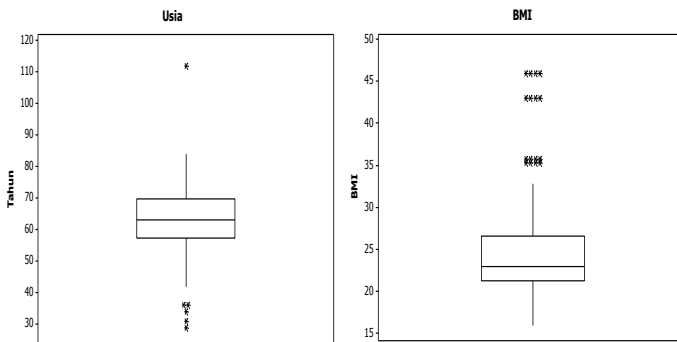
Pada Gambar 4.2 menunjukkan tekanan darah pasien DM tipe II dimana dari 344 pasien DM tipe II, sebanyak 228 pasien memiliki tekanan darah 130 mmHg, 63 pasien memiliki tekanan darah 140 mmHg, 36 pasien memiliki tekanan darah 120 mmHg, 11 pasien memiliki tekanan darah 135 mmHg 6 pasien memiliki tekanan darah yang tinggi, yaitu 145 mmHg.

Tabel 4.4 Tabulasi kadar glukosa darah (Y) dan senam diabetes (X_5)

| Pasien DM Tipe II | Tidak mengikuti | Mengikuti |
|----------------------------|-----------------|-----------|
| | Senam | Senam |
| Kadar Glukosa darah normal | 20 | 147 |
| Kadar Glukosa darah tinggi | 26 | 151 |
| Total | 46 | 298 |

Aktivitas fisik yang digunakan dalam analisis adalah keikutsertaan pasien dalam mengikuti senam diabetes yang dilaksanakan klinik As-Shafa pada hari minggu setiap 2 minggu sekali. Pada Tabel 4.4 yaitu tabel tabulasi antara kadar glukosa darah (Y) dan keikutsertaan senam diabetes (X_5) diketahui bahwa dari 344 pasien, 46 pasien yang tidak mengikuti senam diabetes, dimana 20 pasien memiliki kadar glukosa darah normal dan 26 memiliki kadar glukosa darah tinggi. Selanjutnya pada Gambar 4.3, yaitu *boxplot* variabel usia (X_2) dan BMI (X_3) dapat diketahui bahwa pada variabel usia (X_2) memiliki nilai kuartil pertama (Q_1) sebesar 57, median sebesar 63, kuartil ketiga (Q_3) sebe-

sar 70 serta terdapat 6 data *outlier* dan terlihat bahwa data untuk variabel usia (X_2) cenderung menjulur ke kiri (negatif), sedangkan untuk variabel BMI (X_3) memiliki nilai kuartil pertama (Q_1) sebesar 21,360, median sebesar 23,050, kuartil ketiga (Q_3) sebesar 26,610 serta terdapat 10 data *outlier* dan terlihat bahwa data untuk variabel BMI (X_3) cenderung menjulur ke kanan (positif).



Gambar 4.3 Boxplot variabel usia (X_2) dan BMI (X_3)

4.2 Pembentukan Model MARS

Tahap dalam pembentukan model MARS dilakukan dengan kombinasi antara jumlah maksimum Basis Fungsi (BF), Maksimum Interaksi (MI) dan Minimum Observasi (MO) diantara model knot. MI merupakan banyak interaksi yang dapat terjadi dalam model. Jika MI yang digunakan adalah 1, berarti tidak ada interaksi antar variabel dalam model. Jika MI yang digunakan 2, maka ada interaksi antar 2 variabel dalam model. Begitu pula untuk MI yang digunakan 3, maka ada interaksi antar 3 variabel dalam model. Variabel prediktor yang digunakan dalam penelitian ini ada sebanyak lima variabel, sehingga banyaknya BF yang digunakan adalah sebanyak 10, 15 dan 20. Untuk nilai MO yang digunakan ialah 0,1, 2, 3, 5 dan 10, sehingga banyaknya model yang mungkin berdasarkan kombinasi BF, MI dan MO sebanyak 54 model. Data yang digunakan pada pembentukan model MARS adalah data *training* yang berjumlah 259

data dan hasil pemodelan MARS untuk data diabetes melitus tipe II di Klinik As-Shafa, Sidoarjo pada penelitian ini ditunjukkan pada Tabel 4.5

Tabel 4.5 Hasil Pemodelan MARS

| Model | BF | MI | MO | GCV | MSE | R² | 1-APER |
|--------------|-----------|-----------|-----------|--------------|--------------|----------------------|---------------|
| 1 | 10 | 1 | 0 | 0,243 | 0,238 | 0,547 | 0,525 |
| 2 | 10 | 1 | 1 | 0,243 | 0,238 | 0,547 | 0,525 |
| 3 | 10 | 1 | 2 | 0,243 | 0,238 | 0,548 | 0,525 |
| 4 | 10 | 1 | 3 | 0,243 | 0,238 | 0,548 | 0,525 |
| 5 | 10 | 1 | 5 | 0,243 | 0,238 | 0,547 | 0,525 |
| 6 | 10 | 1 | 10 | 0,243 | 0,238 | 0,548 | 0,525 |
| 7 | 10 | 2 | 0 | 0,245 | 0,238 | 0,547 | 0,525 |
| 8 | 10 | 2 | 1 | 0,245 | 0,238 | 0,547 | 0,525 |
| 9 | 10 | 2 | 2 | 0,245 | 0,238 | 0,547 | 0,525 |
| 10 | 10 | 2 | 3 | 0,245 | 0,238 | 0,547 | 0,525 |
| 11 | 10 | 2 | 5 | 0,245 | 0,238 | 0,547 | 0,525 |
| 12 | 10 | 2 | 10 | 0,245 | 0,238 | 0,547 | 0,525 |
| 13 | 10 | 3 | 0 | 0,245 | 0,238 | 0,547 | 0,525 |
| 14 | 10 | 3 | 1 | 0,245 | 0,238 | 0,547 | 0,525 |
| 15 | 10 | 3 | 2 | 0,245 | 0,238 | 0,547 | 0,525 |
| 16 | 10 | 3 | 3 | 0,245 | 0,238 | 0,547 | 0,525 |
| 17 | 10 | 3 | 5 | 0,245 | 0,238 | 0,547 | 0,525 |
| 18* | 10 | 3 | 10 | 0,251 | 0,232 | 0,562 | 0,587 |
| 19 | 15 | 1 | 0 | 0,243 | 0,238 | 0,548 | 0,525 |
| 20 | 15 | 1 | 1 | 0,243 | 0,238 | 0,547 | 0,525 |
| 21 | 15 | 1 | 2 | 0,243 | 0,238 | 0,548 | 0,525 |
| 22 | 15 | 1 | 3 | 0,243 | 0,238 | 0,548 | 0,525 |
| 23 | 15 | 1 | 5 | 0,243 | 0,238 | 0,548 | 0,525 |
| 24 | 15 | 1 | 10 | 0,243 | 0,238 | 0,548 | 0,525 |
| 25 | 15 | 2 | 0 | 0,245 | 0,238 | 0,548 | 0,525 |
| 26 | 15 | 2 | 1 | 0,245 | 0,238 | 0,548 | 0,525 |
| 27 | 15 | 2 | 2 | 0,245 | 0,238 | 0,548 | 0,525 |
| 28 | 15 | 2 | 3 | 0,245 | 0,238 | 0,548 | 0,525 |
| 29 | 15 | 2 | 5 | 0,245 | 0,238 | 0,548 | 0,525 |
| 30 | 15 | 2 | 10 | 0,245 | 0,238 | 0,548 | 0,525 |
| 31 | 15 | 3 | 0 | 0,245 | 0,238 | 0,547 | 0,525 |
| 32 | 15 | 3 | 1 | 0,245 | 0,238 | 0,547 | 0,525 |
| 33 | 15 | 3 | 2 | 0,245 | 0,238 | 0,547 | 0,525 |

Tabel 4.5 Hasil Pemodelan MARS (lanjutan)

| Model | BF | MI | MO | GCV | MSE | R² | 1-APER |
|--------------|-----------|-----------|-----------|------------|------------|----------------------|---------------|
| 34 | 15 | 3 | 3 | 0,245 | 0,238 | 0,547 | 0,525 |
| 35 | 15 | 3 | 5 | 0,245 | 0,238 | 0,547 | 0,525 |
| 36 | 15 | 3 | 10 | 0,251 | 0,232 | 0,562 | 0,58,7 |
| 37 | 20 | 1 | 0 | 0,243 | 0,238 | 0,548 | 0,525 |
| 38 | 20 | 1 | 1 | 0,243 | 0,238 | 0,547 | 0,525 |
| 39 | 20 | 1 | 2 | 0,243 | 0,238 | 0,548 | 0,525 |
| 40 | 20 | 1 | 3 | 0,243 | 0,238 | 0,548 | 0,525 |
| 41 | 20 | 1 | 5 | 0,243 | 0,238 | 0,548 | 0,525 |
| 42 | 20 | 1 | 10 | 0,243 | 0,238 | 0,548 | 0,525 |
| 43 | 20 | 2 | 0 | 0,254 | 0,238 | 0,548 | 0,525 |
| 44 | 20 | 2 | 1 | 0,254 | 0,238 | 0,548 | 0,525 |
| 45 | 20 | 2 | 2 | 0,254 | 0,238 | 0,548 | 0,525 |
| 46 | 20 | 2 | 3 | 0,254 | 0,238 | 0,548 | 0,525 |
| 47 | 20 | 2 | 5 | 0,254 | 0,238 | 0,548 | 0,525 |
| 48 | 20 | 2 | 10 | 0,254 | 0,238 | 0,548 | 0,525 |
| 49 | 20 | 3 | 0 | 0,245 | 0,238 | 0,547 | 0,525 |
| 50 | 20 | 3 | 1 | 0,245 | 0,238 | 0,547 | 0,525 |
| 51 | 20 | 3 | 2 | 0,245 | 0,238 | 0,547 | 0,525 |
| 52 | 20 | 3 | 3 | 0,245 | 0,238 | 0,547 | 0,525 |
| 53 | 20 | 3 | 5 | 0,245 | 0,238 | 0,547 | 0,525 |
| 54 | 20 | 3 | 10 | 0 | 0,251 | 0,521 | 0,521 |

* Model terpilih

Tabel 4.3 merupakan hasil pemodelan MARS untuk diabetes melitus tipe II di Klinik As-Shafa, Sidoarjo dengan lima variabel yang diduga mempengaruhinya. Dari semua kemungkinan model berdasarkan kombinasi nilai BF, MI dan MO, maka didapatkan model terbaik MARS dengan kriteria memiliki nilai GCV dan MSE terkecil serta nilai R² dan akurasi terbesar, yaitu model ke-18 dengan nilai BF=10, MI=3 dan MO=10 yang menghasilkan nilai GCV=0,251, MSE=0,232, R²=0,562 yang artinya model ke-18 mampu menjelaskan variabilitas variabel prediktor sebanyak 56,2%, sedangkan 43,8% lainnya dijelaskan oleh variabel lain diluar model dan 1-APER =0,587 artinya presentase model yang dipilih untuk dapat tepat mengklasifikasikan suatu data adalah sebesar

58,7% dan semakin tinggi nilai 1-APER menunjukkan model baik untuk klasifikasi. Berdasarkan model ke-18, maka Model MARS yang didapatkan yaitu sebagai berikut.

$$\hat{f}(x) = 0,627 + 0,002 * BF_5 - 0,027 * BF_6 - 0,002 * BF_8 \quad (4.1)$$

dengan,

$$BF_1 = \max(0, X_1 - 59,00)$$

$$BF_3 = (X_5 = 0) * BF_1$$

$$BF_5 = \max(0, X_3 - 15,980) * BF_3$$

$$BF_6 = (X_2 = 0) * BF_1$$

$$BF_7 = (X_2 = 1) * BF_1$$

$$BF_8 = \max(0, X_3 - 15,980) * BF_7$$

Dapat dilihat bahwa pada persamaan (4.1) setiap basis fungsi memuat interaksi antar variabel prediktor dengan rincian sebagai berikut :

BF_5 : memuat interaksi antara variabel *Body mass index* (X_3), aktivitas olahraga (X_5) serta Usia (X_1) pasien DM tipe II

BF_6 : memuat interaksi antara variabel jenis kelamin (X_2) dan Usia (X_1) pasien DM tipe II

BF_8 : memuat interaksi antara variabel *Body mass index* (X_3), jenis kelamin (X_2) serta Usia (X_1) pasien DM tipe II

Berdasarkan model yang telah didapat pada persamaan (4.1), maka dapat dihitung nilai probabilitas pasien memiliki kadar glukosa darah tinggi ($\hat{\pi}(x)$) dan probabilitas pasien memiliki kadar glukosa darah normal ($1 - \hat{\pi}(x)$) adalah sebagai berikut.

$$\hat{\pi}(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}} \quad (4.2)$$

$$(1 - \hat{\pi}(x)) = 1 - \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}} \quad (4.3)$$

dengan,

$$e^{\hat{f}(x)} = \exp(0,627 + 0,002(1) - 0,027(1) - 0,002(1)) = e^{0,6} = 1,8221$$

Sehingga,

$$\hat{\pi}(x) = \frac{e^{\hat{f}(x)}}{1 + e^{\hat{f}(x)}} = \frac{1,8221}{1 + 1,8221} = 0,6456$$

$$(1 - \hat{\pi}(x)) = 1 - 0,6456 = 0,3544$$

Jadi probabilitas pasien DM tipe II di klinik As-Shafa memiliki kadar glukosa darah tinggi adalah 0,6456 dan probabilitas pasien DM tipe II di klinik As-Shafa memiliki kadar glukosa darah normal adalah 0,3544.

Interpretasi model MARS berdasarkan persamaan (4.1) dan probabilitas pasien memiliki kadar glukosa darah tinggi pada persamaan (4.2) serta probabilitas pasien memiliki kadar glukosa darah normal pada persamaan (4.3) adalah sebagai berikut.

$$\begin{aligned} 1. \quad BF_5 &= \max(0, X_3 - 15,980) * BF_3 \\ BF_3 &= (X_5 = 0) * BF_1 \\ BF_1 &= \max(0, X_1 - 59,00) \\ \hat{\pi}(x) &= \frac{e^{(0,627+0,002(1))}}{1 + e^{(0,627+0,002(1))}} = \frac{1,8757}{1 + 1,8757} = 0,652 \\ (1 - \hat{\pi}(x)) &= 1 - 0,652 = 0,348 \end{aligned}$$

Probabilitas pasien DM tipe II di klinik As-Shafa dengan nilai BMI lebih dari 15,980 dan tidak mengikuti senam diabetes serta usia lebih dari 59 untuk memiliki kadar glukosa darah tinggi sebesar 0,652 dan probabilitas untuk memiliki kadar glukosa darah normal sebesar 0,348 saat nilai untuk BF yang lain konstan. Hal tersebut sangat sesuai Tabel 4.4, dimana pasien yang tidak mengikuti senam diabetes lebih banyak memiliki kadar glukosa darah tinggi daripada kadar glukosa darah normal.

$$\begin{aligned} 2. \quad BF_6 &= (X_2 = 0) * BF_1 \\ BF_1 &= \max(0, X_1 - 59,00) \\ \hat{\pi}(x) &= \frac{e^{(0,627-0,027(1))}}{1 + e^{(0,627-0,027(1))}} = \frac{1,8221}{1 + 1,8221} = 0,646 \\ (1 - \hat{\pi}(x)) &= 1 - 0,646 = 0,354 \end{aligned}$$

Probabilitas pasien DM tipe II di klinik As-Shafa yang berjenis kelamin perempuan dan memiliki usia lebih dari 59 untuk memiliki kadar glukosa darah tinggi sebesar 0,646 dan probabilitas untuk memiliki kadar glukosa darah normal

sebesar 0,354 saat nilai untuk BF yang lain konstan. Interpretasi BF_6 tersebut sesuai dengan Tabel 4.2 dimana perempuan lebih banyak yang memiliki kadar glukosa darah tinggi dan pada BF berpengaruh signifikan ketika berinteraksi dengan variabel usia saja.

$$3. \quad BF_8 = \max(0, X_3 - 15,980) * BF_7$$

$$BF_7 = (X_2 = 1) * BF_1$$

$$BF_1 = \max(0, X_1 - 59,00)$$

$$\hat{\pi}(x) = \frac{e^{(0,627-0,002(1))}}{1 + e^{(0,627-0,002(1))}} = \frac{1,8682}{1 + 1,8682} = 0,651$$

$$(1 - \hat{\pi}(x)) = 1 - 0,651 = 0,349$$

Probabilitas pasien DM tipe II di klinik As-Shafa dengan nilai BMI lebih dari 15,980 yang berjenis kelamin laki-laki dan memiliki usia lebih dari 59 untuk memiliki kadar glukosa darah tinggi sebesar 0,651 dan probabilitas untuk memiliki kadar glukosa darah normal sebesar 0,349 saat nilai untuk BF yang lain konstan. Hal tersebut sesuai dengan Tabel 4.2 dimana laki-laki lebih banyak yang memiliki kadar glukosa darah normal yang berpengaruh signifikan dalam model ketika berinteraksi dengan 2 variabel, yaitu usia dan BMI.

Selanjutnya dari model MARS pada persamaan (4.1) dapat diketahui bahwa terdapat empat variabel prediktor yang masuk ke dalam model, yaitu variabel usia (X_1), jenis kelamin (X_2), BMI (X_3) dan aktivitas olahraga (X_5). Untuk melihat sejauh mana variabel-variabel tersebut mempengaruhi pembentukan model MARS, maka dapat dilihat pada tabel *variable importance* pada lampiran B.

Tabel 4.6 Tingkat kepentingan variabel prediktor

| Variabel | Tingkat kepentingan | -gcv |
|------------------------------|---------------------|-------|
| Jenis Kelamin (X_2) | 100,000 % | 0,257 |
| BMI (X_3) | 60,165 % | 0,253 |
| Usia (X_1) | 42,682 % | 0,252 |
| Aktivitas Olahraga (X_5) | 34,576 % | 0,251 |
| Tekanan Darah (X_4) | 0,000 % | 0,251 |

Berdasarkan Tabel 4.6 dapat diketahui dimana variabel jenis kelamin (X_2) adalah variabel terpenting pada model MARS dengan tingkat kepentingan 100%, kemudian diikuti oleh variabel BMI (X_3) dengan tingkat kepentingan 60,165%. Pada urutan ketiga yaitu variabel usia (X_1) dengan tingkat kepentingan 42,682%, serta urutan keempat yaitu variabel aktivitas olahraga (X_5) dengan tingkat kepentingan 34,576%, sedangkan variabel tekanan darah (X_4) tidak memiliki tingkat kepentingan (0%) karena sudah terwakili oleh empat variabel sebelumnya. Nilai $-gcv$ menunjukkan seberapa besar pengurangan nilai gcv jika variabel tersebut dikeluarkan dalam model MARS.

Dari model MARS terbaik, maka dapat dilakukan pengujian signifikansi parameter

1. Pengujian regresi simultan

Hipotesis :

$$H_0 : \gamma_5 = \gamma_6 = \gamma_8 = 0 \text{ (model tidak signifikan)}$$

$$H_1 : \text{minimal terdapat satu } \gamma_m \neq 0 ; m = 5, 6, 8 \text{ (model signifikan)}$$

Taraf signifikan : 5%

Statistik uji :

Tabel 4.7 Hasil pengujian secara simultan

| F_{hitung} | P_{value} |
|---------------------------|--------------------------|
| 7,911 | 0,000045 |

Berdasarkan Tabel 4.7, dengan daerah kritis ialah tolak H_0 jika nilai $F_{hitung} > F_{0,05(3;340)}$ atau nilai $P\text{-value} < 0,05$, maka diperoleh kesimpulan tolak H_0 karena nilai $F_{hitung} (7,911) > F_{tabel} (2,63)$ yang artinya model signifikan, sehingga dapat digunakan untuk memodelkan kasus diabetes melitus tipe II.

2. Pengujian regresi parsial

Hipotesis :

$$H_0 : \gamma_m = 0 \text{ (koefisien } \gamma_m \text{ tidak berpengaruh terhadap model)}$$

$$H_1 : \gamma_m \neq 0 \text{ untuk setiap } m, \text{ dimana } m = 5, 6, 8$$

Taraf signifikan : 5%

Statistik uji :

Tabel 4.8 Hasil pengujian secara parsial

| Parameter | Estimasi | S.E | t_{hitung} | P-value |
|-----------|----------|-------|--------------|---------|
| Konstan | 0,627 | 0,039 | 15,916 | 0,000 |
| BF5 | 0,002 | 0,007 | 2,946 | 0,004 |
| BF6 | -0,027 | 0,007 | -3,760 | 0,002 |
| BF8 | -0,002 | 0,004 | -3,782 | 0,001 |

Hasil pengujian secara parsial pada Tabel 4.8 diatas menunjukkan bahwa koefisien α_5, α_6 dan α_8 berpengaruh signifikan terhadap, sehingga dapat digunakan untuk memodelkan kasus diabetes melitus tipe II, nilai $|t\text{-ratio}|$ untuk semua koefisien $> t_{tabel}$ (2,25).

4.3 Ketetapan Klasifikasi MARS

Dari hasil model MARS terbaik yang diperoleh, maka akan dilakukan klasifikasikan untuk mengetahui seberapa baik model tersebut berdasarkan nilai 1-APER, sensitivitas, spesifisitas serta nilai G-means berdasarkan tabel klasifikasi yang diperoleh dari data *training* pada Tabel 4.9 dan data *testing* pada Tabel 4.8.

Tabel 4.9 Klasifikasi DM tipe II data *training*

| Data aktual | Data prediksi | | Total aktual |
|-----------------------|---------------|-----|--------------|
| | 0 | 1 | |
| 0 | 57 | 67 | 124 |
| 1 | 36 | 99 | 135 |
| Total prediksi | 93 | 166 | 259 |

Berdasarkan Tabel 4.9 diketahui bahwa dari 259 data *testing*, 57 tepat diklasifikasikan ke dalam kategori tekanan darah rendah dan 67 salah diklasifikasikan dari kategori tekanan darah rendah ke dalam kategori tekanan darah tinggi, sedangkan 99 data tepat diklasifikasikan ke dalam kategori tekanan darah tinggi dan 36 lainnya salah diklasifikasikan dari kategori tekanan darah tinggi ke dalam kategori tekanan darah rendah. Selanjutnya dari model MARS terbaik dilakukan klasifikasi data *testing* yang hasilnya dapat dilihat pada Tabel 4.10. Tabel 10 menunjukkan bahwa dari 85 data *testing*, 17 tepat diklasifikasikan ke dalam kategori tekanan darah rendah dan 26 salah diklasifikasikan dari kategori

tekanan darah rendah ke dalam kategori tekanan darah tinggi, sedangkan 28 tepat diklasifikasi ke dalam kategori tekanan darah tinggi dan 14 data salah diklasifikasikan dari kategori tekanan darah tinggi ke dalam kategori tekanan darah rendah

Tabel 4.10 Klasifikasi DM tipe II data *testing*

| Data aktual | Data prediksi | | Total aktual |
|-----------------------|---------------|----|--------------|
| | 0 | 1 | |
| 0 | 17 | 26 | 43 |
| 1 | 14 | 28 | 42 |
| Total prediksi | 31 | 54 | 85 |

Berdasarkan informasi dari Tabel 4.9 dan 4.10, maka dapat dihitung nilai ketetapan klasifikasinya, yaitu 1-APER dan G-means pada Tabel 4.11.

Tabel 4.11 Ketetapan klasifikasi MARS

| Nilai ketetapan Klasifikasi | Data <i>training</i> | Data <i>testing</i> |
|--------------------------------|----------------------|---------------------|
| 1-APER | 0,602 | 0,529 |
| G-means | 1,092 | 1,031 |

Dari Tabel 4.11 dapat diketahui bahwa nilai 1-APER untuk data *training* sebesar 0,602, sedangkan untuk data *testing* 0,529. Nilai G-means menunjukkan bahwa metode MARS mampu memprediksi data tiap kelas dengan baik. Nilai G-means data *training* lebih tinggi daripada data *testing*, sehingga dapat dikatakan model kurang baik untuk mengklasifikasikan data baru. Model dikatakan baik jika dapat mengklasifikasikan data *testing* dengan baik, namun pada permasalahan ini kurang sesuai, hal ini mungkin dikarenakan pembagian data *testing* dan *training* untuk setiap kategori kurang tepat.

4.4 Ketetapan klasifikasi *Boosting* MARS

Dengan harapan dapat menurunkan tingkat kesalahan klasifikasi dari model MARS, maka dilakukan *resampling* pada hasil pemodelan MARS menggunakan metode *boosting*. Langkah yang dilakukan adalah melakukan *boosting* dari model MARS dengan

berbagai kombinasi pengulangan. Selanjutnya dilakukan klasifikasi pada masing-masing kombinasi pengulangan pada data *training* dan data *testing* yang dapat dilihat pada Tabel 4.12 dan 4.13. Berdasarkan Tabel 4.12 dapat diketahui nilai 1-APER, serta G-means dari masing-masing kombinasi pengulangan, dimana pada data *training* dengan kombinasi pengulangan 250 dan seterusnya memiliki nilai 1-APER serta G-means yang sama. Hal ini menandakan saat kombinasi pengulangan 250, hasil klasifikasi sudah stabil, sehingga saat kombinasi pengulangan ditambah, hasilnya akan sama. Jika dibandingkan dengan 1-APER metode MARS pada data *training*, maka *boosting* MARS memiliki nilai ketetapan klasifikasi lebih tinggi, sehingga dapat dikatakan bahwa metode *boosting* dapat meningkatkan tingkat ketetapan klasifikasi dari MARS.

Tabel 4.12 Perbandingan Ketetapan Klasifikasi MARS dan *Boosting* MARS Data *Training*

| Metode | Kombinasi pengulangan | 1-APER | G-means |
|-----------------|-----------------------|--------------|--------------|
| MARS | - | 0,602 | 1,092 |
| | 50 | 0,683 | 1,057 |
| | 100 | 0,683 | 1,057 |
| | 250* | 0,722 | 1,107 |
| Boosting | 500 | 0,722 | 1,107 |
| MARS | 600 | 0,722 | 1,107 |
| | 700 | 0,722 | 1,107 |
| | 800 | 0,722 | 1,107 |
| | 900 | 0,722 | 1,107 |
| | 1000 | 0,722 | 1,107 |

* Kombinasi pengulangan stabil

Selanjutnya ketetapan klasifikasi *Boosting* MARS untuk data *testing* yang dapat dilihat pada Tabel 4.13 yang menunjukkan data *testing* dengan kombinasi pengulangan 800 dan seterusnya data baru stabil. Jika dibandingkan dengan metode MARS pada data *testing*, maka *boosting* MARS memiliki nilai 1-APER serta G-means yang lebih tinggi. Sesuai dengan tujuan *boosting* MARS yaitu meningkatkan nilai 1-APER (akurasi), maka terlihat bahwa

nilai 1-APER dari data *training* dan *testing* mengalami peningkatan dan fokus *boosting* yaitu pada sel klasifikasi yang mengalami kesalahan klasifikasi (APER), sehingga terlihat nilai APER dari metode *boosting* lebih kecil daripada MARS, maka dapat dikatakan bahwa metode *boosting* dapat meningkatkan tingkat ketetapan klasifikasi MARS pada kasus DM tipe II di klinik As-Shafa, Sidoarjo

Tabel 4.13 Perbandingan Ketetapan Klasifikasi MARS dan *Boosting* MARS Data *Testing*

| Metode | Kombinasi pengulangan | 1-APER | G-means |
|-------------------------|-----------------------|--------------|--------------|
| <i>Boosting</i> MARS | MARS | | |
| | - | 0,529 | 1,031 |
| | 50 | 0,671 | 1,130 |
| | 100 | 0,671 | 1,130 |
| | 250 | 0,682 | 1,143 |
| | 500 | 0,694 | 1,148 |
| | 600 | 0,694 | 1,148 |
| | 700 | 0,718 | 1,180 |
| | 800* | 0,741 | 1,204 |
| | 900 | 0,741 | 1,204 |
| | 1000 | 0,741 | 1,204 |

*Kombinasi pengulangan stabil

BAB V

KESIMPULAN DAN SARAN

Pada bab ini dijelaskan hasil analisis secara lebih singkat, yaitu kesimpulan. Selain itu, dalam bab ini juga akan diberikan saran berdasarkan hasil analisis yang telah dilakukan.

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut.

1. Model terbaik yang menggambarkan hubungan antara faktor-faktor terjadinya DM tipe II di klinik As-Shafa Sidoarjo dengan menggunakan model MARS adalah sebagai berikut.

$$\hat{f}(x) = 0,627 + 0,002 * BF_5 - 0,027 * BF_6 - 0,002 * BF_8$$

dengan,

$$BF_1 = \max(0, X_1 - 59,00)$$

$$BF_3 = (X_5 = 0) * BF_1$$

$$BF_5 = \max(0, X_3 - 15,980) * BF_3$$

$$BF_6 = (X_2 = 0) * BF_1$$

$$BF_7 = (X_2 = 1) * BF_1$$

$$BF_8 = \max(0, X_3 - 15,980) * BF_7$$

Pada kombinasi BF=10, MI=3 dan MO=10 didapatkan empat variabel yang berkontribusi dalam model, yaitu variabel jenis kelamin (X_2), variabel BMI (X_3), variabel usia (X_1) dan variabel aktivitas olahraga (X_5) dengan tingkat kepentingan masing-masing 100%, 60,165%, 42,682% dan 34,576% serta berdasarkan hasil pengujian parameter, diketahui bahwa model telah memenuhi uji simultan dan parsial, sehingga model dapat digunakan untuk kasus diabetes melitus tipe II.

2. Nilai 1-APER untuk data *training* sebesar 0,602 dan untuk data *testing* 0,529, sedangkan nilai G-means data *testing* lebih rendah daripada data *training*, sehingga dapat dikatakan model kurang untuk mengklasifikasikan data baru. Model dikatakan baik jika dapat mengklasifikasikan data *testing* dengan baik, namun pada permasalahan ini kurang sesuai, hal ini mungkin dikarenakan pembagian data *testing* dan *training* untuk setiap kategori kurang tepat

3. Dengan menggunakan metode *boosting* MARS pada data *training* dengan kombinasi pengulangan 250, hasil klasifikasi sudah stabil dengan nilai 1-APER sebesar 0,722 dan G-means sebesar 1,107 sedangkan untuk data *testing* pada kombinasi pengulangan 800, hasil klasifikasi baru stabil dengan nilai 1-APER sebesar 0,741 dan G-means sebesar 1,204, sehingga saat kombinasi pengulangan ditambah, maka akan memiliki hasil yang sama dan dapat dikatakan bahwa metode *boosting* dapat meningkatkan tingkat ketetapan klasifikasi MARS.

5.2 Saran

Dalam penelitian ini hanya menggunakan 5 (lima) variabel umum DM tipe II yang hasilnya variabel yang digunakan hanya mampu menjelaskan 56,2% variabel respon, sehingga penelitian lebih lanjut perlu dilakukan penambahan lebih banyak variabel yang digunakan.

DAFTAR PUSTAKA

- Adityaningrum, A. (2016). *Estimasi Propensity Score Matching Menggunakan Pendekatan Multivariate Adaptive Regression Splines (Tesis)*. Surabaya: ITS.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis Second Edition*. New Jersey: John Wiley & Son, Inc.
- BPJS. (2014). *Program Rujuk Balik Bagi Peserta JKN*. Jakarta.
- Depkes R.I. (2014). *Pusat Data dan Informasi Kementerian Kesehatan RI*. Jakarta Selatan: <http://www.depkes.go.id/resources/download/pusdatin/infodatin/infodatin-diabetes.pdf>.
- Depkes RI. (2005). *Pharmaceutical Care untuk Penyakit Diabetes Melitus*. Jakarta.
- Draper, N., & Smith, H. (1992). *Analisis Regresi Terapan*. (B. Sumantri, Trans.) Jakarta: PT Gramedia Pustaka Utama.
- Erniati. (2013). *Faktor-faktor yang berhubungan dengan Diabetes Melitus tipe 2 pada Lanjt Usia di Pos Pembinaan Terpadu Keurahan Cempaka Putih tahun 2012 (Skripsi)*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.
- Eubank, R. (1999). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.Inc.
- Fernanda, J. W. (2012). *Boosting Neural Network dan Boosting CART pada klasifikasi Diabetes Melitus Tipe II*. Surabaya: Program Pasca Sarjana : Institut Teknologi Sepuluh Nopember.
- Friedman, J. H. (1990). *Multivariate Adaptive Regression Spline*. California: Stanford University.
- Handayani, S. A. (2003). *Faktor-faktor Risiko Diabetes Melitus Tipe-2 di Semarang dan Sekitarnya (Tesis)*. Semarang: Universitas Diponegoro.
- Hestie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning*. California: Stanford.

- Hidayanti, A. A. (2014). *Boosting Multivariate Adaptive Regression Spline (MARS) Binary Respinse untuk Klasifikasi Kemiskinan di Kabupaten Jombang*. Institut Teknologi Sepuluh Nopember, Statistika, Surabaya.
- Indonesia, P. E. (2015). *Konsensus Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 di Indonesia*. Jakarta: Perkumpulan Endokrinologi Indonesia.
- Johnson, R. A., & Winchern, D. W. (1992). *Applied Multivariate Statistical Analysis*. United States: Prentice Hall.
- Ningtyas, D. W. (2013). *Analisis Kualitas Hidup Pasien Diabetes Melitus Tipe II di RSUD Bangil Kabupaten Pasuruan*. Jember: Universitas Jember.
- Otok, B. W. (2010). Pendekatan Multivariate Adaptive Regression Spline (MARS) pada Pengelompokkan Zona Musim Suatu Wilayah. *Journal Statistik Vol 10. FMIPA ITS Surabaya*.
- Permatasari, E. O. (2013). *Pendekatan Boosting Multivariate Adaptive Regression Spline(Boosting MARS) untuk Klasifikasi Kemiskinan di Provinsi Jawa Timur (Tesis)*. Surabaya: ITS .
- Wahba, G. (1990). *Spline Models for Observational Data*. University of Winconsin at Madison.
- Wu, X., & Kumar, V. (2008). *Top 10 algorithms in data mining*. Boca Raton: Taylor & Francis Group.

LAMPIRAN

Lampiran A. Data Rekap Medis Pasien DM tipe II

1. Data training

| No | y | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ |
|-----|---|----------------|----------------|----------------|----------------|----------------|
| 1 | 1 | 75 | 1 | 23,42 | 130 | 1 |
| 2 | 0 | 67 | 1 | 25,71 | 130 | 0 |
| 3 | 1 | 76 | 1 | 42,97 | 130 | 0 |
| 4 | 0 | 59 | 1 | 20,57 | 130 | 1 |
| 5 | 1 | 82 | 1 | 18,90 | 130 | 1 |
| 6 | 1 | 42 | 0 | 26,84 | 130 | 1 |
| 7 | 0 | 49 | 0 | 26,17 | 140 | 0 |
| 8 | 0 | 71 | 0 | 20,69 | 140 | 1 |
| 9 | 1 | 63 | 1 | 22,66 | 140 | 1 |
| 10 | 1 | 57 | 0 | 21,91 | 135 | 1 |
| 11 | 1 | 53 | 1 | 25,11 | 130 | 1 |
| 12 | 0 | 75 | 1 | 22,15 | 130 | 1 |
| 13 | 1 | 62 | 1 | 23,07 | 140 | 1 |
| 14 | 0 | 34 | 1 | 24,68 | 130 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 250 | 1 | 57 | 0 | 32,03 | 130 | 1 |
| 251 | 0 | 63 | 1 | 25,40 | 130 | 1 |
| 252 | 1 | 70 | 0 | 23,05 | 130 | 1 |
| 253 | 0 | 63 | 1 | 45,92 | 120 | 1 |
| 254 | 0 | 55 | 0 | 23,59 | 145 | 1 |
| 255 | 0 | 74 | 1 | 28,91 | 130 | 1 |
| 256 | 0 | 58 | 1 | 20,69 | 140 | 1 |
| 257 | 1 | 49 | 0 | 21,63 | 130 | 1 |
| 258 | 0 | 62 | 0 | 27,77 | 135 | 0 |
| 259 | 1 | 66 | 1 | 21,09 | 120 | 1 |

2. Data Testing

| No | y | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ |
|----|---|----------------|----------------|----------------|----------------|----------------|
| 1 | 1 | 50 | 1 | 21,09 | 140 | 1 |
| 2 | 1 | 58 | 0 | 35,76 | 130 | 1 |
| 3 | 0 | 62 | 1 | 15,98 | 120 | 1 |
| 4 | 1 | 63 | 0 | 22,51 | 130 | 1 |
| 5 | 0 | 75 | 1 | 22,31 | 140 | 1 |
| 6 | 1 | 57 | 0 | 28,95 | 130 | 1 |
| 7 | 0 | 61 | 0 | 21,09 | 130 | 0 |
| 8 | 0 | 72 | 1 | 29,90 | 130 | 0 |
| 9 | 1 | 74 | 1 | 21,63 | 130 | 1 |
| 10 | 0 | 78 | 1 | 26,77 | 130 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 77 | 0 | 63 | 1 | 25,40 | 130 | 1 |
| 78 | 1 | 70 | 0 | 23,05 | 130 | 1 |
| 79 | 0 | 63 | 1 | 45,92 | 120 | 1 |
| 80 | 0 | 55 | 0 | 23,59 | 145 | 1 |
| 81 | 0 | 74 | 1 | 28,91 | 130 | 1 |
| 82 | 0 | 58 | 1 | 20,69 | 140 | 1 |
| 83 | 1 | 49 | 0 | 21,63 | 130 | 1 |
| 84 | 0 | 62 | 0 | 27,77 | 135 | 0 |
| 85 | 1 | 66 | 1 | 21,09 | 120 | 1 |

Keterangan :

Y = kadar glukosa darah (1= tinggi, 0=normal)

X₁ = Usia

X₂ = Jenis kelamin (1=laki-laki, 0=perempuan)

X₃ = *Body Mass Index*

X₄ = Tekanan darah

X₅ = Aktivitas olahraga

X₆ = Komplikasikesehatan

Lampiran B. Output Model Terbaik MARS (BF=10, MI=3 dan MO=10)

LEARNING SAMPLE STATISTICS

=====

| VARIABLE | MEAN | SD | N | SUM |
|----------|---------|-------|---------|-----------|
| Y | 0.521 | 0.501 | 259.000 | 135.000 |
| X1 | 63.201 | 9.140 | 259.000 | 16369.000 |
| X2 | 0.413 | 0.493 | 259.000 | 107.000 |
| X3 | 24.511 | 4.876 | 259.000 | 6348.410 |
| X4 | 131.158 | 5.940 | 259.000 | 33970.000 |
| X5 | 0.865 | 0.343 | 259.000 | 224.000 |

AUTOMATIC LEVEL SETTINGS

| NAME | LEVELS | MINIMUM |
|------|--------|---------|
| X2 | 2 | 0 |
| X5 | 2 | 0 |

Ordinal Response

| | min | Q25 | Q50 | Q75 | max |
|---|-------|-------|-------|-------|-------|
| Y | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |

Ordinal Predictor Variables: 3

| | min | Q25 | Q50 | Q75 | max |
|----|---------|---------|---------|---------|---------|
| X1 | 31.000 | 57.000 | 63.000 | 69.000 | 112.000 |
| X3 | 15.980 | 21.360 | 23.050 | 26.610 | 45.920 |
| X4 | 120.000 | 130.000 | 130.000 | 130.000 | 145.000 |

Categorical Predictor Variables: 2

| | Variable | NLEV | Actual | | | Internal | | Counts |
|---|----------|------|--------|----|---|----------|-----|--------|
| 2 | X2 | 2 | 0. | 1. | 1 | 2 | 152 | 107 |
| 5 | X5 | 2 | 0. | 1. | 1 | 2 | 35 | 224 |

Forward Stepwise Knot Placement

=====

| BasFn(s) | GCV | IndBsFns | EfPrms | V | variable | Knot | Parent | BsF |
|----------|-------|----------|--------|-----|----------|------|--------|-----|
| 0 | 0.251 | 0.0 | 1.0 | | | | | |
| 2 | 1 | 0.246 | 2.0 | 6.0 | X1 | | 59.000 | |

| | | | | | | | | |
|---|----|---|-------|-----|------|----|---------|----|
| 1 | 4 | 3 | 0.252 | 3.0 | 10.0 | X5 | 10 | X1 |
| 3 | | 5 | 0.257 | 4.0 | 14.0 | X3 | 15.980 | X5 |
| 1 | 7 | 6 | 0.263 | 5.0 | 18.0 | X2 | 10 | X1 |
| 7 | | 8 | 0.270 | 6.0 | 22.0 | X3 | 15.980 | X2 |
| 3 | | 9 | 0.276 | 7.0 | 26.0 | X4 | 120.000 | X5 |
| 4 | 10 | | 0.284 | 8.0 | 30.0 | X2 | 10 | X5 |

Final Model (After Backward Stepwise Elimination)

| Basis | Fun | Coefficient | Variable | Parent | Knot |
|-------|-----|-------------|----------|--------|--------|
| 0 | | 0.627 | | | |
| 5 | | 0.002 | X3 | X5 | 15.980 |
| 6 | | -0.027 | X2 | X1 | |
| 8 | | -0.002 | X3 | X2 | 15.980 |

Piecewise Linear GCV = 0.251, #efprms = 11.875

ANOVA Decomposition on 3 Basis Functions

| fun | std. dev. | -gcv | #bsfns | #efprms | variable |
|-----|-----------|-------|--------|---------|----------|
| 1 | 0.120 | 0.257 | 1 | 3.625 | X1 X2 |
| 2 | 0.094 | 0.252 | 1 | 3.625 | X1 X3 |
| | | | | | X5 |
| 3 | 0.124 | 0.257 | 1 | 3.625 | X1 X2 |
| | | | | | X3 |

Piecewise Cubic Fit on 3 Basis Functions, GCV = 0.254

Relative Variable Importance

| Variable | Importance | -gcv |
|----------|------------|---------|
| 2 | X2 | 100.000 |
| 3 | X3 | 60.165 |
| 5 | X5 | 42.682 |
| 1 | X1 | 34.576 |
| 4 | X4 | 0.000 |

ORDINARY LEAST SQUARES RESULTS

```

=====
N: 259.000                                R-SQUARED: 0.085
MEAN DEP VAR: 0.521                      ADJ R-SQUARED: 0.074
                                           UNCENTERED R-SQUARED = R-0 SQUARED: 0.562

```

| PARAMETER | ESTIMATE | S.E. | T-RATIO | P-VALUE |
|------------------|----------|-------------|---------|-------------|
| Constant | 0.627 | 0.039 | 15.916 | .999201E-15 |
| Basis Function 5 | 0.002 | .733042E-03 | 2.946 | 0.004 |
| Basis Function 6 | -0.027 | 0.007 | -3.760 | .211098E-03 |
| Basis Function 8 | -0.002 | .476570E-03 | -3.782 | .193792E-03 |

```

-----
F-STATISTIC = 7.911                      S.E. OF REGRESSION = 0.482
P-VALUE = .458441E-04                   RESIDUAL SUM OF SQUARES = 59.130
[MDF,NDF] = [ 3, 255 ]                   REGRESSION SUM OF SQUARES = 5.503
-----

```

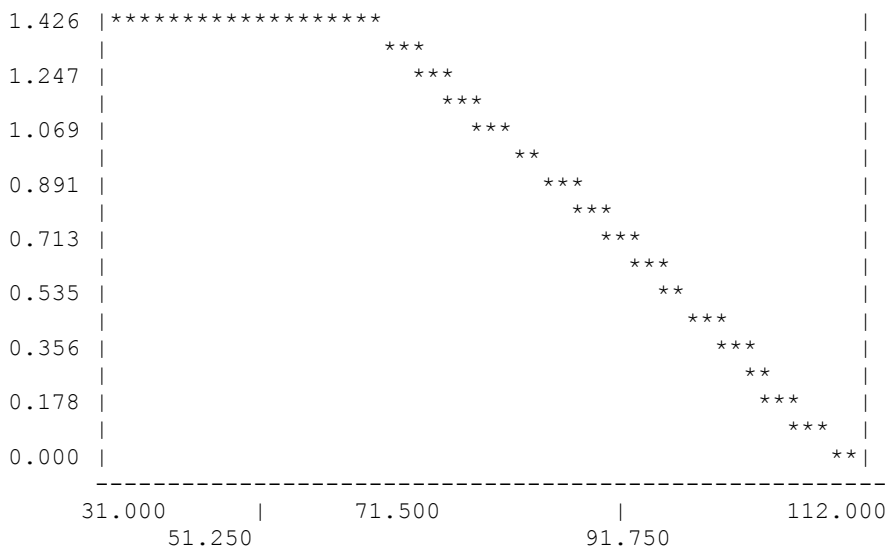
The Following Graphics Are Piecewise Linear

srf 1: x(1), x(3). max = 1.4881

CATEGORICAL - ORDINAL INTERACTION:

X2 = 1 0

CURVE 1: X1 , max = 1.4257



```
srf      2:  x(      1), x(      3).  max =      1.2419

1 curves and 2 surfaces.

Basis Functions
=====
BF1 = max(0, X1 - 59.000);
BF3 = ( X5 = 0) * BF1;
BF5 = max(0, X3 - 15.980) * BF3;
BF6 = ( X2 = 0) * BF1;
BF7 = ( X2 = 1) * BF1;
BF8 = max(0, X3 - 15.980) * BF7;

Y = 0.627 + 0.002 * BF5 - 0.027 * BF6 - 0.002 * BF8;
model Y = BF5 BF6 BF8;

=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

| Actual Class | Predicted Class | | Actual Total |
|-----------------------------------------|-----------------|-----------------|--------------|
| | 0 | 1 | |
| 0 | 19.000 | 105.000 | 124.000 |
| 1 | 2.000 | 133.000 | 135.000 |
| Pred. Tot. | 21.000 | 238.000 | 259.000 |
| Correct | 0.153 | 0.985 | |
| Success Ind. | -0.326 | 0.464 | |
| Tot. Correct | 0.587 | | |
| Sensitivity: | 0.153 | Specificity: | 0.985 |
| False Reference: | 0.095 | False Response: | 0.441 |
| Reference = Class 0, Response = Class 1 | | | |

```
-----
```

Lampiran C. Program *Boosting* MARS

50 kombinasi pengulangan

```
library(RWeka)
WOW(AdaBoostM1)
data<-read.csv('D://modelterbaik.csv', sep=';',header=T)
data1<-read.csv('D://data.csv', sep=';',header=T)
m1<-AdaBoostM1(Y~b0+BF7+BF8+BF10,data=data,
control=Weka_control(I=50))
predict(m1)
table(predict(m1),data1$Y)
summary(m1)
```

100 kombinasi pengulangan

```
library(RWeka)
WOW(AdaBoostM1)
data<-read.csv('D://modelterbaik.csv', sep=';',header=T)
data1<-read.csv('D://data.csv', sep=';',header=T)
m1<-AdaBoostM1(Y~b0+BF7+BF8+BF10,data=data,
control=Weka_control(I=100))
predict(m1)
table(predict(m1),data1$Y)
summary(m1)
```

500 kombinasi pengulangan

```
library(RWeka)
WOW(AdaBoostM1)
data<-read.csv('D://modelterbaik.csv', sep=';',header=T)
data1<-read.csv('D://data.csv', sep=';',header=T)
m1<-AdaBoostM1(Y~b0+BF7+BF8+BF10,data=data,
control=Weka_control(I=500))
predict(m1)
table(predict(m1),data1$Y)
summary(m1)
```

Lampiran D. *Output Boosting* MARS (Data training dengan 250 pengulangan)

```
> predict(m1)
[1] kadar glukosa darah tinggi kadar glukosa darah tinggi
[3] kadar glukosa darah tinggi kadar glukosa darah tinggi
[5] kadar glukosa darah tinggi kadar glukosa darah tinggi
[7] kadar glukosa darah tinggi kadar glukosa darah tinggi
[9] kadar glukosa darah tinggi kadar glukosa darah tinggi
[11] kadar glukosa darah tinggi kadar glukosa darah tinggi
[13] kadar glukosa darah tinggi kadar glukosa darah tinggi
[15] kadar glukosa darah tinggi kadar glukosa darah tinggi
[17] kadar glukosa darah tinggi kadar glukosa darah rendah
[19] kadar glukosa darah tinggi kadar glukosa darah tinggi
[21] kadar glukosa darah tinggi kadar glukosa darah rendah
[23] kadar glukosa darah tinggi kadar glukosa darah tinggi
[25] kadar glukosa darah tinggi kadar glukosa darah tinggi
[27] kadar glukosa darah tinggi kadar glukosa darah tinggi
[29] kadar glukosa darah rendah kadar glukosa darah tinggi
[31] kadar glukosa darah tinggi kadar glukosa darah tinggi
[33] kadar glukosa darah tinggi kadar glukosa darah tinggi
[35] kadar glukosa darah tinggi kadar glukosa darah tinggi
[37] kadar glukosa darah tinggi kadar glukosa darah tinggi
[39] kadar glukosa darah tinggi kadar glukosa darah tinggi
[41] kadar glukosa darah tinggi kadar glukosa darah tinggi
[43] kadar glukosa darah tinggi kadar glukosa darah tinggi
[45] kadar glukosa darah rendah kadar glukosa darah tinggi
[47] kadar glukosa darah tinggi kadar glukosa darah tinggi
[49] kadar glukosa darah tinggi kadar glukosa darah tinggi
[51] kadar glukosa darah tinggi kadar glukosa darah tinggi
[53] kadar glukosa darah tinggi kadar glukosa darah tinggi
[55] kadar glukosa darah tinggi kadar glukosa darah tinggi
[57] kadar glukosa darah tinggi kadar glukosa darah tinggi
[59] kadar glukosa darah tinggi kadar glukosa darah tinggi
[61] kadar glukosa darah tinggi kadar glukosa darah tinggi
[63] kadar glukosa darah tinggi kadar glukosa darah tinggi
[65] kadar glukosa darah tinggi kadar glukosa darah tinggi
[67] kadar glukosa darah tinggi kadar glukosa darah tinggi
[69] kadar glukosa darah tinggi kadar glukosa darah tinggi
[71] kadar glukosa darah tinggi kadar glukosa darah tinggi
[73] kadar glukosa darah tinggi kadar glukosa darah tinggi
[75] kadar glukosa darah tinggi kadar glukosa darah tinggi
[77] kadar glukosa darah tinggi kadar glukosa darah tinggi
[79] kadar glukosa darah tinggi kadar glukosa darah tinggi
[81] kadar glukosa darah tinggi kadar glukosa darah tinggi
[83] kadar glukosa darah tinggi kadar glukosa darah tinggi
[85] kadar glukosa darah tinggi kadar glukosa darah tinggi
[87] kadar glukosa darah tinggi kadar glukosa darah tinggi
[89] kadar glukosa darah tinggi kadar glukosa darah tinggi
```

[illegible]

```

[191] kadar glukosa darah tinggi kadar glukosa darah tinggi
[193] kadar glukosa darah tinggi kadar glukosa darah tinggi
[195] kadar glukosa darah tinggi kadar glukosa darah tinggi
[197] kadar glukosa darah tinggi kadar glukosa darah tinggi
[199] kadar glukosa darah rendah kadar glukosa darah tinggi
[201] kadar glukosa darah tinggi kadar glukosa darah tinggi
[203] kadar glukosa darah rendah kadar glukosa darah tinggi
[205] kadar glukosa darah tinggi kadar glukosa darah tinggi
[207] kadar glukosa darah tinggi kadar glukosa darah tinggi
[209] kadar glukosa darah tinggi kadar glukosa darah tinggi
[211] kadar glukosa darah tinggi kadar glukosa darah tinggi
[213] kadar glukosa darah tinggi kadar glukosa darah tinggi
[215] kadar glukosa darah tinggi kadar glukosa darah tinggi
[217] kadar glukosa darah tinggi kadar glukosa darah tinggi
[219] kadar glukosa darah tinggi kadar glukosa darah tinggi
[221] kadar glukosa darah tinggi kadar glukosa darah tinggi
[223] kadar glukosa darah tinggi kadar glukosa darah tinggi
[225] kadar glukosa darah tinggi kadar glukosa darah tinggi
[227] kadar glukosa darah tinggi kadar glukosa darah tinggi
[229] kadar glukosa darah tinggi kadar glukosa darah tinggi
[231] kadar glukosa darah tinggi kadar glukosa darah tinggi
[233] kadar glukosa darah tinggi kadar glukosa darah tinggi
[235] kadar glukosa darah tinggi kadar glukosa darah tinggi
[237] kadar glukosa darah tinggi kadar glukosa darah tinggi
[239] kadar glukosa darah tinggi kadar glukosa darah tinggi
[241] kadar glukosa darah tinggi kadar glukosa darah tinggi
[243] kadar glukosa darah tinggi kadar glukosa darah tinggi
[245] kadar glukosa darah rendah kadar glukosa darah tinggi
[247] kadar glukosa darah tinggi kadar glukosa darah tinggi
[249] kadar glukosa darah tinggi kadar glukosa darah tinggi
[251] kadar glukosa darah tinggi kadar glukosa darah tinggi
[253] kadar glukosa darah rendah kadar glukosa darah tinggi
[255] kadar glukosa darah rendah kadar glukosa darah tinggi
[257] kadar glukosa darah tinggi kadar glukosa darah tinggi
[259] kadar glukosa darah tinggi

```

```

Levels: kadar glukosa darah rendah kadar glukosa darah
tinggi

```

```

> table(predict(m1),data1$Y)

```

| | kadar glukosa darah rendah |
|----------------------------|----------------------------|
| kadar glukosa darah rendah | 10 |
| kadar glukosa darah tinggi | 117 |

| | kadar glukosa darah tinggi |
|----------------------------|----------------------------|
| kadar glukosa darah rendah | 11 |
| kadar glukosa darah tinggi | 121 |

```

> summary(m1)

```

```

=== Summary ===

```



```

Correctly Classified Instances      187
72.2008 %
Incorrectly Classified Instances    72
27.7992 %
Kappa statistic                     0.2721
Mean absolute error                 0.3795
Root mean squared error            0.4202
Relative absolute error            82.3926 %
Root relative squared error        87.5874 %
Coverage of cases (0.95 level)     100 %
Mean rel. region size (0.95 level) 100 %
Total Number of Instances          259

=== Confusion Matrix ===

  a    b  <-- classified as
21  72 |   a = kadar glukosa darah rendah
 0 166 |   b = kadar glukosa darah tinggi

```

Lampiran E. Pembuktian Pengujian Dependensi

Pengujian dilakukan untuk mengetahui hubungan antara variabel kadar glukosa darah (Y) dengan variabel prediktor jenis kelamin (X_2) dan variabel aktivitas olahraga (X_5).

Hipotesis :

H_0 : Tidak ada hubungan antara variabel respon dengan variabel prediktor

H_1 : ada hubungan antara variabel respon dengan variabel prediktor

Taraf Signifikan :5%

| Variabel | χ^2 | P_value | Keputusan | Kesimpulan |
|----------|----------|---------|-------------------|------------------------------|
| X_2 | 2,396 | 0,075 | Tolak H_0 | Berpengaruh signifikan |
| X_5 | 0,546 | 0,281 | Gagal tolak H_0 | Tidak berpengaruh signifikan |

Berdasarkan hasil pengujian diketahui bahwa dari 2 variabel prediktor yang diujikan dependensi hanya variabel jenis kelamin (X_2) yang berpengaruh signifikan, sedangkan variabel aktivitas olahraga (X_5) tidak berpengaruh signifikan.

Output SPSS pengujian Dependensi**Chi-Square Tests JK**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---------------------------------------|--------------------|----|--------------------------|-------------------------|-------------------------|
| Pearson Chi-Square | 2.396 ^a | 1 | .122 | | |
| Continuity Correction ^b | 2.069 | 1 | .150 | | |
| Likelihood Ratio | 2.398 | 1 | .122 | | |
| Fisher's Exact Test | | | | .127 | .075 |
| Linear-by-Linear Association | 2.389 | 1 | .122 | | |
| N of Valid Cases | 344 | | | | |

Chi-Square Tests Aktivitas Olahraga

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---------------------------------------|-------------------|----|--------------------------|-------------------------|-------------------------|
| Pearson Chi-Square | .546 ^a | 1 | .460 | | |
| Continuity Correction ^b | .337 | 1 | .562 | | |
| Likelihood Ratio | .548 | 1 | .459 | | |
| Fisher's Exact Test | | | | .527 | .281 |
| Linear-by-Linear Association | .544 | 1 | .461 | | |
| N of Valid Cases | 344 | | | | |

Lampiran F. Legalitas Data**SURAT KETERANGAN**

Saya yang bertanda tangan dibawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMIPA – ITS dengan identitas berikut :

Nama : Widya Umroatun Sa'diyah

NRP : 1315105027

Telah mengambil data di instansi/perusahaan kami :

Nama Instansi : Klinik "As-Shafa"

Divisi/Bagian : Laboratorium/Penyakit Kronik (DM)

Sejak tanggal 1 Maret 2017 sampai dengan 1 Juni 2017 untuk keperluan Tugas Akhir/Thesis semester ~~Gasal~~/Genap* 2016/2017

2. Tidak Keberatan/~~Keberatan~~* nama perusahaan dicantumkan dalam Tugas Akhir/Thesis mahasiswa Statistika yang akan di dimpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/~~Keberatan~~* bahwa hasil analisis data dari perusahaan dipublikasikan dalam *E-journal* ITS yaitu Jurnal Sains dan Seni ITS.

Surabaya, 6 Juni 2017

KLINIK "AS-SHAFa"



**KLINIK UTAMA
"AS SHAFa"**
Jl. Legan Sutoyo No. 11 Medang, Waru - Sidoarjo
Telp. (031) 8559001 - 8559002

dr. Nurul Hidayati

Penanggung Jawab Klinik

*(coret yang tidak perlu)

BIODATA PENULIS



Penulis dengan nama lengkap Widya Umroatun Sa'diyah lahir di Kabupaten Lamongan, 03 Mei 1994. Penulis merupakan anak pertama dari pasangan Nawi dan Sumarsih. Penulis mempunyai seorang adik perempuan yang bernama Mirna Halimatus Sa'diyah yang sedang menempuh Sekolah dasar di SDN 1 Wedoro, Sukorame, Lamongan.. Jenjang pendidikan formal yang ditempuh penulis adalah TK Hidup Santosa Lamongan pada tahun 2000-

2001, SDN Wedoro, kecamatan Sukorame, Lamongan pada tahun 2001-2006, SMP Negeri Sukorame, Lamongan pada tahun 2006-2009, SMA Negeri Ploso, Jombang pada tahun 2009-2012, Diploma III Statistika FMIPA ITS Surabaya pada tahun 2012-2015, kemudian melanjutkan Lintas Jalur S-1 Statistika FMIPA ITS dengan NRP 1315105027. Apabila pembaca ingin berdiskusi tentang Tugas Akhir ini dapat menghubungi penulis melalui email widyaumroatun@gmail.com

(Halaman Ini Sengaja Dikosongkan)